统计模拟（第4版）

Simulation

“……本书内容丰富，不论作为教材还是参考书都非常值得推荐。”

——美国统计学报

“本书是一本非常优秀的教材，强调了计算机在模拟技术上的应用。阐述清晰，例题富于趣味性。”

——亚马逊网上书店评论

“《统计模拟》很好地做到了与计算技术发展的同步。据我所知，这是统计模拟方面唯一的一本教材。它的内容如此出色，可以说是无可替代的。”

—— Kris Ostaszewski（伊利诺伊大学）

统计模拟是一门新兴的统计学和计算机结合的学科，因其便利性和经济性而广泛应用于统计学、数学、精算科学、工程学、物理学等众多领域，用以获得精确而有效的解决方案。

本书是国际知名统计学家 Sheldon M. Ross 所著的经典教材，已被加州大学伯克利分校、哥伦比亚大学等多所名校采用。书中涵盖了统计模拟的最新方法和技术，提供了丰富的实例，备受业界推崇。



Sheldon M. Ross 国际知名概率与统计学家，南加州大学工业工程与运筹系系主任。毕业于斯坦福大学统计系，曾在加州大学伯克利分校任教多年。研究领域包括：随机模型、仿真模拟、统计分析、金融数学等。Ross 教授著述颇丰，他的多种畅销数学和统计教材均产生了世界性的影响，如 *Introduction to Probability Models*（《应用随机过程：概率模型导论》）、*A First Course in Probability*（《概率论基础教程》）等（均由人民邮电出版社出版）。

本书译自原版 *Simulation*，并由 Elsevier 授权出版。



本书相关信息请访问：
图灵网站 http://www.turingbook.com
读者/作者热线：(010) 88593802
反馈/投稿/推荐信箱：contact@turingbook.com

分类建议 数学/统计

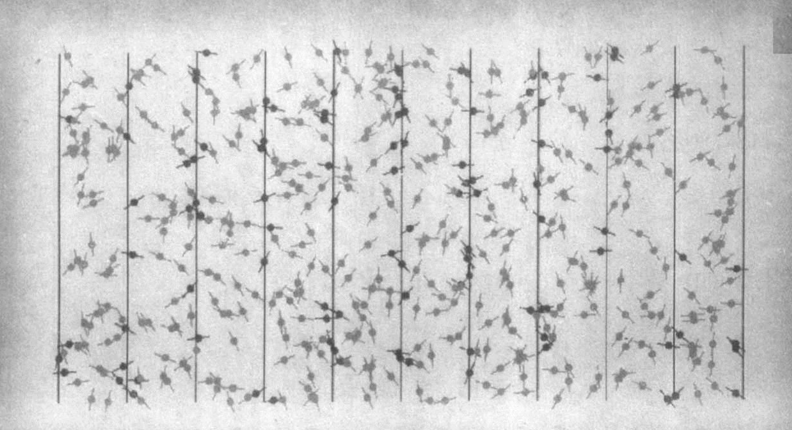
人民邮电出版社网址 www.ptpress.com.cn



ISBN 978-7-115-16085-0/O1

定价：39.00 元

TURING 图灵数学·统计学丛书 11



Simulation

统计模拟

（第4版）

Sheldon M. Ross 著

陈广雷 邹长亮 译

人民邮电出版社

北 京

Simulation

统计模拟

（第 4 版）

[美] Sheldon M. Ross 著

王兆军 陈广雷 邹长亮 译



图书在版编目（CIP）数据

统计模拟：第4版/（美）罗斯，（Ross, S. M）著；王兆军，陈广雷，邹长亮译.
—北京：人民邮电出版社，2007.7
（图灵数学·统计学丛书）

ISBN 978-7-115-16085-0

Ⅰ. 统… Ⅱ. ①罗…②王…③陈…④邹… Ⅲ. 统计－模拟实验－高等学校－教材
Ⅳ. C8-33

中国版本图书馆CIP数据核字（2007）第052005号

内 容 提 要

统计模拟是数理统计中非常有用的工具之一，它是利用计算机产生某概率模型的随机数，再通过这些随机数来模拟真实模型．本书首先介绍了产生某些分布随机数的一些方法，之后又较详细地介绍了统计模拟中常用的一些方法，如离散事件模拟方法、方差缩减技术、模拟数据的统计分析方法、统计验证方法、MCMC 方法等；并通过某些实例，对这些方法的应用进行了较详细的说明．本书最后还提供了不同难度的习题．

本书可作为高等院校数学、统计学、科学计算、保险学和精算学等专业的本科教材，也可作为工程技术人员和精算师等应用工作者的参考用书．

图灵数学·统计学丛书

统计模拟（第4版）

◆ 著　　［美］Sheldon M. Ross
　译　　王兆军　陈广雷　邹长亮
　责任编辑　明永玲　王　利

◆ 人民邮电出版社出版发行　　北京市崇文区夕照寺街14号
　邮编　100061　　电子函件　315@ptpress.com.cn
　网址　http://www.ptpress.com.cn
　北京铭成印刷有限公司印刷
　新华书店总店北京发行所经销

◆ 开本：700×1000　1/16
　印张：16.75
　字数：344千字　　　2007年7月第1版
　印数：1－4 000册　　2007年7月北京第1次印刷

著作权合同登记号　图字：01-2007-1485号

ISBN 978-7-115-16085-0/O1

定价：39.00元

读者服务热线：(010)88593802　印装质量热线：(010)67129223

版权声明

Simulation, *4th Edition* by Sheldon M. Ross, ISBN: 0-12-598063-9.

Authorized Simplified Chinese translation edition published by the Proprietor.

ISBN: 978-981-259-847-9.

Elsevier (Singapore) Pte Ltd.

3 Killiney Road,

#08-01 Winsland House I

Singapore 239519

Tel: (65)6349-0200

Fax: (65)6733-1817

First Published 2007

2007 年初版

译者简介

王兆军 南开大学教授、博士生导师. 现任南开大学数学科学学院统计学系系主任、中国概率统计学会理事、中国现场统计研究会理事、天津数学会秘书长. 1987年毕业于南开大学，获理学学士学位；1990年毕业于华东师范大学，获理学硕士学位；1995年于南开大学获理学博士学位.

陈广雷 中国人民武装警察部队学院副教授. 1992年毕业于东北师范大学，获理学学士学位；2000年毕业于北京师范大学，获理学硕士学位；现为南开大学在读博士生.

邹长亮 2003年毕业于南开大学，获理学学士学位；2006年毕业于南开大学，获理学硕士学位；2006年9月起在南开大学攻读博士学位.

译 者 序

统计模拟方法不仅是统计专业的一门重要课程，而且也是越来越多的理工学院、商学院、经济学院、医学院、农学院部分专业中的选修课程. 虽然国内目前关于统计计算的教材已有若干本，但却没有一本书专门系统介绍统计模拟，尤其是最近几年发展起来且得到广泛应用的新方法，如 Bootstrap(自助法)、MCMC(马氏链蒙特卡洛法) 和模拟退火算法等. 然而，这本由 Sheldon M. Ross 编写的《统计模拟》教材，恰好是这方面的补充与完善.

本书首先介绍了随机数的产生方法，之后又给出了几种产生连续与离散随机变量的方法，并通过某些例子简述了某些模拟方法及模拟数据的统计分析方法. 然后，提出了若干方差缩减技术，如对偶变量法、控制变量法、分层抽样法、重要抽样法、公共变量法等，并介绍了几种模拟检验方法，如 χ^2 拟合优度检验、科尔莫戈罗夫 – 斯米尔诺夫检验等. 在介绍 MCMC 方法的一章，又讲述了 Hasting-Metropolis 算法、吉布斯抽样、模拟退火和重要抽样方法等. 最后又对随机模拟的其他一些专题进行了讨论.

本书在讲述方法的同时，还注重这些方法在金融、优化等方面的应用，且给出了较丰富的参考文献. 虽然本书内容很丰富，但因其所需的概率统计知识相对较少，所以很适合作为本科生自学或教材用书，而且其中某些内容也可供统计专业的、研究生参考学习.

我们很高兴能有机会将该书推荐给国内的读者，也非常感谢人民邮电出版社图灵公司的编辑在本书出版过程中给予我们的大力支持和帮助. 另外, 也要感谢几位没有署名的审稿者.

本书的翻译工作是由 3 名作者合作完成的，其中第 1 章至第 6 章由王兆军翻译，第 7 章、第 8 章由陈广雷翻译，第 9 章至第 11 章由邹长亮翻译，全书由王兆军统校.

由于译者无论是英文、中文水平还是统计专业知识都很有限，译文中难免会有不妥之处，欢迎广大读者批评指正.

译 者

2006 年 12 月于南开园

前　　言

概述

在用随机模型去描述实际现象时，人们既要考虑所选模型的逼真性，又要考虑所选模型的可分析性. 即是说，一个无法进行数学分析的模型，即使它能很好地反映实际问题，也是毫无用处. 于是，人们转而把注意力放在考虑一个能近似反映实际问题的模型. 然而，随着高性能且价格相对便宜的计算机的出现，产生了一种用计算机尽可能真实地模拟随机现象，之后再利用模拟结果去分析它的新方法.

本书将展示如何利用随机模拟结果对模型进行分析的方法. 特别地，本书将给出如何利用计算机产生随机数（或称伪随机数）的方法，讨论如何利用这些随机数来产生任一分布的随机变量的值，阐述如何利用离散随机事件的概念去模拟一个依时间而变化的随机模型的行为，以及给出通过模拟一个随机系统的行为而得到某感兴趣参数的估计. 在模拟过程中，我们将考虑何时停止模拟及如何给定所考虑的估计的置信度等问题. 本书将给出几种能改进常用模拟估计的模拟方法. 另外，我们还将探讨如何利用随机模拟去判断所用随机模型是否与实际数据相符的问题.

本版的新内容

本版的变化主要体现在以下几个方面.

- 大部分章都增加了新习题.
- 增加关于生成贝努利随机变量序列的新结果 (例 4e).
- 关于选取最优指数分布的筛选法生成伽玛随机变量的新结果 (5.2 节).
- 新增例 8p 给出了求取当所有癌细胞被杀死后仍存活的健康细胞数目的分布. 重写了关于分层抽样的 8.4 节，包括有关事后分层及每层最优模拟次数的材料.
- 新增 8.5 节，此节包括把分层抽样应用于到达为泊松过程的系统分析 (8.5.1 节)、单调函数的多重积分的计算 (8.5.2 节) 和复合随机向量 (8.5.3 节) 三部分.
- 新增了有效利用方差缩减技术计算随机排列和随机子集函数的 8.9 节.

各章内容

第 1 章是绪言，它给出了一个有意思的典型案例. **第 2 章**是概率论简介. 本章自成体系并在假设读者不熟悉概率论的情况下给出简单介绍. **第 3 章**涉及随机数及如何由计算机产生伪随机数. **第 4 章**和**第 5 章**考虑如何利用随机数来产生离散随机变量和连续随机变量.

第 6 章 介绍离散随机事件方法以分析随时间而变化的随机系统, 并给出了多个例子, 如涉及单个或多服务员的排队系统、保险风险模型、仓储问题、机器维修模型及股票期权的执行模型等.

第 7 章 的主要内容是统计, 我们假设本书的一般读者事先没有学过统计. 本章先从非常基本的统计概念开始讲起, 最后以自助法 (Bootstrap) 结束 (自助法在随机模拟中非常有用).

第 8 章 涉及方差缩减技术这一重要内容. 我们将尝试通过寻找一个具有相同均值和较小方差的估计以改进常用的模拟估计. 本章首先介绍对偶变量 (Antithetic variables) 方法. 本章附录的证明显示, 当把此方法用来估计某单调函数的期望时, 它总能减小方差, 并只占用较少的计算机存储空间. 其次, 我们引入控制变量并阐述它在减小方差上的作用, 且用在排队系统、可靠性系统、列表重排序问题和 21 点游戏 (Blackjack) 中的例子说明如何有效地利用控制变量. 我们也将提出如何利用回归软件包承担控制变量的计算量. 之后考虑利用条件期望以减小方差, 并把它应用于估计 π 和分析有限容量的排队系统中. 结合控制变量法, 条件期望法也被用来估计更新过程中某时刻的平均事件数. 作为减小方差的另一工具, 分层抽样被应用于处理带有变化到达率的排队问题和积分求取的例子中. 在估计某视频纸牌游戏平均收益的例子中, 我们解释和说明了条件期望和分层抽样这两种减小方差方法间的关系. 分层抽样在到达为泊松过程的排队系统、多重积分的计算和复合随机向量中的应用. 然后, 我们考虑重要抽样法. 重要抽样方法是一种非常有效的估计小概率的方差缩减方法. 为说明这一点, 我们引入倾斜分布的概念, 说明如何利用它和重要抽样法来估计一个卷积的小尾部概率. 我们还把重要抽样方法应用于排队论、随机游动、随机排列和基于小概率事件的条件期望的计算上. 本章最后的一个方差缩减方法涉及公共随机数流的应用. 8.7 节介绍用多种方差缩减技术相结合为奇异股票期权 (exotic stock option) 定价.

第 9 章 涉及某些统计验证方法. 当有实际数据可用时, 这些统计方法可以用来验证随机模型的有效性. 本章将介绍如 χ^2 检验和科尔莫戈罗夫 - 斯米尔诺夫等拟合优度检验. 本章的其他小节讨论两样本问题、多样本问题和检验一个过程是否为泊松过程的统计假设检验.

第 10 章 涉及马氏链蒙特卡洛 (MCMC) 方法. 这些方法极大地拓展了近几年随机模拟的应用. 对于一个随机向量 $\boldsymbol{X}$, 用随机模拟方法估计 $\theta = E[h(\boldsymbol{X})]$ 的标准步骤, 是先模拟产生多个与 $\boldsymbol{X}$ 独立同分布的随机变量, 之后用基于这些随机变量的 $h(\boldsymbol{X})$ 的均值作为 θ 的估计. 这就是通常所说的原始估计, 它可以用第 8 章中的方差缩减技术进行改进. 然而, 要使用这种方法, 就必须知道 $\boldsymbol{X}$ 的分布且能模拟此分布. 但正如本章的许多例子, 虽然我们知道 $\boldsymbol{X}$ 的分布, 但我们却不能直接模拟

此随机向量 $\boldsymbol{X}$; 又如本章的其他例子, 虽然其分布可以写成若干因子的乘积, 但并不完全已知. 对于这两种情况, 上述通常估计 θ 的方法均不可行. 但是近几年, 一个基于生成一个马氏链的新方法在随机模拟中得到了广泛的应用, 此马氏链的极限分布同 $\boldsymbol{X}$ 的分布, 且用函数 h 在此马氏链的各状态点的平均值估计 θ. 本章将研究 MCMC 方法. 我们首先在 10.2 节引入马氏链并给出它的某些性质. 10.3 节介绍 Hastings-Metropolis 算法, 它是用来生成一个极限分布为给定因子乘积的马氏链的常用方法. 一种产生来自一 "组合" 集之随机元素的方法也在 10.3 节给出. 10.4 节介绍的吉布斯抽样是最常用的 Hastings-Metropolis 算法. 本节讨论的例子有: 在某区域内生成任两点距离均大于给定值的随机点, 乘积排队网络的分析, 用来预测某些棒球选手本垒打次数的分层贝叶斯统计模型的分析, 以及基于所有试验结果都至少出现一次的条件多项分布的随机模拟等. 10.5 节介绍用于确定性优化问题的模拟退火方法及一个旅行商问题的例子. 本章最后一节介绍重要重抽样算法, 它是第 4 章和第 5 章的筛选 (Acceptance-Rejection) 抽样法的推广, 本节也考虑此算法在贝叶斯统计中的应用.

第 11 章 涉及随机模拟的其他一些专题. 我们将在 11.1 节中熟悉别名方法, 如不考虑计算时间, 它是产生离散随机数的非常有效的方法. 11.2 节讨论二维泊松过程的模拟问题. 我们在 11.3 节给出一个关于相关伯努利随机变量和的协方差的等式, 并指出如何利用此等式估计某些小概率, 且估计方差非常小. 本节还考虑应用此等式去估计系统可靠性 (此估计的效率在估计小系统可靠性问题时高于现有的其他的估计), 去估计指定模式在给定时间发生的概率等问题. 11.4 节考虑马氏链首达时的均值和分布的模拟估计, 以及二维正态变量的尾部概率的计算. 在上述内容的基础上, 11.5 节给出了服从指定马氏链平稳分布的随机变量的模拟方法.

致谢

我们十分感激 Yontha Ath(加州大学长滩分校)、David Butler (俄勒冈州立大学)、Matt Carlton (加利福尼亚州立理工大学)、James Daniel (得克萨斯大学奥斯汀分校)、William Frye (鲍尔州立大学)、Mark Glickman (波士顿大学)、Chuanshu Ji (北卡大学)、Yonghee Kim-Park (加州大学长滩分校)、Donald E. Miller (圣玛丽学院)、Krzysztof Ostaszewski (伊利诺伊州立大学)、Bernardo Pagnocelli, Erol Peköz (波士顿大学)、Yuval Peres (加州大学伯克利分校) 和 Esther Portnoy (伊利诺伊大学厄巴纳 – 尚佩恩分校) 等, 他们提出了十分有益的建议. 我们还要谢谢那些希望匿名的本书的审稿者.

目　录

第1章　绪　　言

考虑如下案例. 一个药剂师想建立一个自己填写处方的小药店. 他计划早九点开门，并希望下午五点前平均每天（周一至周五期间）有 32 个处方需要他填写. 经验显示，他填写一张处方所需的时间是一个均值和标准差分别为 10 分钟和 4 分钟的随机变量. 他假设下午五点后不再收新处方；但若需要，他将留在药店把一天中没有完成的处方写好. 对于此案例，这个药剂师感兴趣的问题可能是：

1. 平均来看，他晚上几点可以离开药店？

2. 晚上五点半时他仍在药店工作的天数的比例是多少？

3. 他填写一张处方所需的平均时间是多少？（要考虑到填写新处方时需等待他写完前面所有的处方）

4. 30 分钟之内能完成的处方比例是多少？

5. 如果他把早九点至晚五点期间接受所有处方的策略做一改变，只有当手中需要填写的处方少于 5 个时才接受新处方，则平均每天会流失多少处方？

6. 上述策略的改变将如何影响问题 1 ~ 4 的答案？

为了用数学方法分析此案例并回答上述问题，我们需要先建立一个概率模型. 为此，我们有必要对上述案例做某些合理的确切的假设. 比如，为了用一概率模型描述平均每天 32 位顾客的到来时间，我们必须对此做些假设. 一个可能的假设是一天中单位时间内（比如每小时）到来的顾客人数在概率意义下是一个常数，尽管另一个更接近实际的假设是顾客到来的人数与时间有关. 然后，我们必须指定一个概率分布（均值为 10，标准差为 4）来刻画填写一个处方的所需时间，以及做一些假设以保证一个处方所需的服务时间服从这个分布或作为其他随机变量的函数（如等待服务处方的数目或时间）. 这就是说，我们必须对顾客到来时间和所需服务时间做某些概率假设，也必须判断用来描述一天的概率分布是依赖于这一天是星期几，还是个与时间无关的常数. 当这些或其他的假设被指定后，我们就可以构造此案例的概率模型了.

一旦概率模型确定后，就可以从理论上定量地给出上述问题的答案. 然而，在实际中我们很难定量地给出这些问题的答案，于是，我们一般通过随机模拟来给出答案. 随机模拟就是通过计算机编程产生随机数以模拟此模型在很多天内的可能结果，并且利用统计理论估计上述问题的答案. 换句话说，通过计算机编程产生用以描述顾客到来时间和服务所需时间的具有特定概率分布的随机变量的值，并由此

得到多天的关于上述问题的随机变量的值，之后利用统计方法估计它们. 例如，如果 1000 天的随机模拟结果显示药剂师在五点半仍在工作的有 122 天，则我们估计问题 2 的答案为 0.122.

为了能完成上述分析，人们必须懂得某些概率论知识，如概率分布、随机变量的独立性等，因此第 2 章为概率论简介. 由于随机数是随机模拟研究的基础，故第 3 章将讨论随机数及如何由计算机产生随机数. 第 4、5 章给出产生服从任一分布的随机数的方法（第 4、5 章分别考虑离散与连续分布的随机数）. 学完第 5 章之后，读者就应有能力对实际问题建立概率模型，并知道如何产生此模型相应随机变量的随机数. 由于有些系统是依时间连续变化的，故第 6 章将考虑如何把上述产生随机数的方法应用于一个实际系统. 在第 6 章中，我们将提出“离散随机事件”的概念，并基于它而得到模拟一系统的系统化方法. 离散随机事件模拟方法即是通过计算机编程（可用读者熟悉的任一种计算机语言）多次模拟此系统而使问题得以解决. 验证此程序可行与否的某些步骤也将在第 6 章给出. 在利用模拟输出结果解答所关心的模型的问题时必须用到某些统计理论，这些将在第 7 章介绍. 第 7 章将首先介绍一些最简单实用的统计概念，之后再引入最近发展起来的在随机模拟中非常有用的“自助法 (Bootstrap)”. 统计研究表明，由随机模拟得到的估计值的方差是衡量随机模拟效率的一个重要指标. 估计值的方差越小，就表明为得到某一给定精度所需的模拟次数越少. 第 8 章将给出一些改进模拟估计方差的方法. 由于方差减小可以提高模拟效率，故减小方差是随机模拟研究中非常重要的课题. 当有实际数据可用时，如何利用随机模拟的结果去验证概率模型与实际情况的接近程度的方法将在第 9 章给出. 第 10 章介绍马氏链蒙特卡洛方法 (MCMC). 近些年，MCMC 方法极大地拓广了随机模拟的研究领域. 第 11 章研究其他一些主题.

2

习　题

1. 下面数据是一个单服务员系统中前 13 个顾客的到达时间和所需的服务时间. 当服务员空闲时，顾客到达后立即得到服务；否则，就要排队等候. 当服务员完成一位顾客的服务后，则立即服务下一个等候时间最长的顾客.

到达时间:	12	31	63	95	99	154	198	221	304	346	411	455	537
服务时间:	40	32	55	48	18	50	47	18	28	54	40	72	12

   (a) 求这 13 位顾客的离开时间.

   (b) 如果有两个服务员且每一服务员可以服务任一顾客，求这 13 位顾客的离开时间.

   (c) 如果假设服务员完成一位顾客的服务后，对等候时间最短的顾客进行服务，求这 13 位顾客的离开时间.

2. 考虑一个先来先服务的服务站. 分别以 A_n, S_n 和 D_n 记顾客 n 的到达时间、所需服务时间和离开时间. 假设此服务站只有一名服务员且开始时站内无顾客. 3

(a) 当 $D_0 = 0$ 时, 请验证

$$D_n - S_n = \max\{A_n, D_{n-1}\}, \quad \forall\, n > 0$$

(b) 当有两名服务员时, 请给出相应的递推公式.

(c) 当有 k 名服务员时, 请给出相应的递推公式.

(d) 请编写求取离开时间的计算机程序（离开时间是到达时间和所需服务时间的函数）, 并验证练习 1 中 (a) 和 (b) 的答案. 4

第2章 概率基础

2.1 样本空间和事件

对于一个事先不知确切结果的试验，其所有可能试验结果构成的集合称为样本空间，记为 S. 例如，考虑一个编号为 1 至 7 的七匹马的赛马试验，其样本空间为

$$S = \{(1, 2, 3, 4, 5, 6, 7)\text{的所有排列}\}$$

其中 $(3, 4, 1, 7, 6, 5, 2)$ 表示 3 号马跑第一，4 号马跑第二，等等.

样本空间的任一子集 A 称为一个事件，即事件是一个包含试验的某些可能结果的集合. 如果一个试验结果包含在 A 中，则我们说事件 A 发生或出现. 例如，在上例中，如果

$$A = \{S\text{中第一个数为5的试验结果}\}$$

则 A 是 5 号马跑第一的事件.

对于任何两个事件 A 和 B，我们定义一个称为二者并的新事件 $A \cup B$，它包含事件 A 或 B 中的所有试验结果. 同样，我们定义一个称为二者交的新事件 AB，它仅包含既在事件 A 又在事件 B 中的所有试验结果. 这就是说，如事件 A 或 B 发生，则事件 $A \cup B$ 发生；如事件 A 和 B 同时发生，则事件 AB 发生. 我们同样可以定义多个事件的并与交. 即，事件 $A_1, \cdots, A_n$ 的并 $\cup_{i=1}^{n} A_i$，被定义为包含在任一 A_i 中的所有试验结果的事件；事件 $A_1, \cdots, A_n$ 的交 $A_1 A_2 \cdots A_n$，被定义为包含在所有 A_i 中的所有试验结果的事件.

5

对于任一事件 A，我们定义其补事件为在样本空间 S 中但不在 A 中的所有试验结果的集合，记为 A^{c}. 事件 A^{c} 发生当且仅当事件 A 不发生. 因为试验结果肯定在样本空间 S 中，故 S^{c} 不包含任何试验结果且它不会发生. 我们称 S^{c} 为空集且记为 $\emptyset$. 如果 $AB = \emptyset$，则 A 与 B 不会同时发生（其原因为二者没有相同试验结果），我们称之为互斥.

2.2 概率公理

假设对于一个样本空间为 S 的试验中的任一事件 A，存在一个常数 $P(A)$，满足下面三个公理：

公理 1 $0 \leqslant P(A) \leqslant 1$.

公理 2 $P(S) = 1$.

公理 3 对任一列互斥事件$A_1, A_2, \cdots$,有

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i),\ n = 1, 2, \cdots, \infty$$

则称$P(A)$为事件 A 的概率.

公理 1 指出事件 A 中任一试验结果出现的概率介于 0 与 1 之间; 公理 2 指出样本空间中的试验结果以概率 1 出现; 公理 3 指出任意多个互斥事件, 至少发生其中之一的概率等于各事件概率之和.

由此三个公理, 概率的许多结论可以得到证明. 比如, 由于 A 和 A^{c} 总是互斥的, 且 $A \cup A^{\mathrm{c}} = S$, 于是由公理 2 和公理 3 有

$$1 = P(S) = P(A \cup A^{\mathrm{c}}) = P(A) + P(A^{\mathrm{c}})$$

或等价地有

$$P(A^{\mathrm{c}}) = 1 - P(A)$$

也可以说, 一个事件不发生的概率等于 1 减去此事件发生的概率. 6

2.3 条件概率和独立性

考虑掷一枚硬币两次的试验, 因为每次试验结果不是正面就是反面, 故此试验的样本空间可以写成

$$S = \{(\mathrm{H,H}), (\mathrm{H,T}), (\mathrm{T,H}), (\mathrm{T,T})\}$$

其中 (H,T) 表示第一次掷出现正面, 第二次掷出反面. 现假设四种结果的出现是等可能的, 则每一结果出现的概率为 1/4. 如果已知第一次试验的结果为正面, 则在此信息下, 两次结果均为正面的概率是多少? 为求此概率, 我们推理如下: 当给定第一次结果为正面时, 试验结果最多有两种情况: (H,H) 或 (H,T). 由于这两种结果在原两次投掷试验中发生的概率相同, 故在此试验中我们可以假设二者发生的概率相同, 即当给定第一次结果为正面时, 上述两种结果 (H,H) 和 (H,T) 发生的条件概率各为 1/2, 另两个结果的条件概率为 0. 因此, 所求的概率即为 1/2.

如果以 A 和 B 分别表示两次投掷均为正面的事件和第一次投掷为正面的事件, 则上面得到的概率即称为 B 发生下 A 的条件概率, 记为

$$P(A|B)$$

对于所有随机试验和任两事件 A 和 B, 我们均可以用上述同样方法得到 $P(A|B)$ 的一般表达式. 也就是说, 事件 B 发生后, 为使事件 A 发生, 则出现的试验结果必须同时属于 A 和 B, 即在 AB 中. 因为 B 已经发生, 故 B 即为新的样本空间. 因此, 事件 $A|B$ 发生的概率等于事件 AB 相对于事件 B 的概率, 即

$$P(A|B)=\frac{P(AB)}{P(B)}$$

为确定事件 A 发生的概率, 人们经常通过另一事件 B 作如下考虑: 求在事件 B 发生条件下 A 的条件概率及在事件 B 不发生条件下 A 的条件概率, 之后注意到

7

$$A=AB\cup AB^{\mathrm{C}}$$

又因为 AB 与 AB^{C} 互斥, 故有

$$P(A)=P(AB)+P(AB^{\mathrm{C}})=P(A|B)P(B)+P(A|B^{\mathrm{C}})P(B^{\mathrm{C}})$$

当我们利用上一公式时, 我们就说我们是利用**事件 B 发生与否的条件**来计算事件 A 的概率 $P(A)$ 的.

例 2a 一保险公司将投保人分成易出险与不易出险两类. 其数据表明, 易出险客户在一年内提出一次索赔的概率为 0.25, 而不易出险客户在一年内提出一次索赔的概率为 0.1. 如果一个新的投保人是易出险客户的概率为 0.4, 则他在一年内索赔一次的概率为多少?

解 以 C 表示一年内索赔一次的事件, B 表示此投保人为易出险客户的事件. 则

$$P(C)=P(C|B)P(B)+P(C|B^{\mathrm{C}})P(B^{\mathrm{C}})=(0.25)(0.4)+(0.1)(0.6)=0.16.$$

□

假设事件 B_i, $i=1,\cdots,n$ 中恰有一个事件发生, 即假设事件 $B_1,B_2,\cdots,B_n$ 互斥且它们的并为样本空间 S. 于是, 我们可以利用 B_i 发生的条件来计算事件 A 的概率, 其原因为

$$A=AS=A\left(\bigcup_{i=1}^{n}B_i\right)=\bigcup_{i=1}^{n}AB_i$$

由此可得

$$P(A)=\sum_{i=1}^{n}P(AB_i)=\sum_{i=1}^{n}P(A|B_i)P(B_i)$$

例 2b 假设某种票证共有 k 种, 且一个新品种的收集独立于前面所收集到的, 而第 j 种票证以概率 p_j 被收集到 ($\sum_{j=1}^{k} p_j = 1$). 求第 n 个被收集到的票证有别于前面的 $n-1$ 个的概率.

8

解 以 N 表示票证 n 为以前没有被收集到的事件. 为利用前面讲的条件概率法计算 $P(N)$, 以 T_j 表示票证 n 是第 j 类票证的事件, 则我们有

$$P(N) = \sum_{j=1}^{k} P(N|T_j)P(T_j) = \sum_{j=1}^{k}(1-p_j)^{n-1}p_j$$

我们注意到 $P(N|T_j) = (1-p_j)^{n-1}$ 的原因为: 当票证 n 为第 j 类票证的条件下, 其为新类型的条件概率等于前 $n-1$ 个票证都不是第 j 类票证的条件概率. □

如上述投掷硬币的例子, 事件 B 发生下 A 的条件概率 $P(A|B)$ 一般并不等于 A 的无条件概率 $P(A)$. 也可以说, 已知事件 B 发生一般会改变事件 A 发生的概率 (如二者互斥, 结论如何?). 当 $P(A|B)$ 等于 $P(A)$ 时, 我们称A 和 B 独立. 由 $P(A|B) = P(AB)/P(B)$ 可知, 如果

$$P(AB) = P(A)P(B)$$

则 A 独立于 B. 又因为此关系关于 A 与 B 是对称的, 故只要 A 独立于 B, 则 B 也独立于 A.

2.4 随机变量

有时, 我们关心的可能是某些试验结果的某一函数的取值, 这些由某些试验结果所确定的量称为随机变量.

一个随机变量 X 的累积分布函数, 或简称分布函数 F 定义为

$$F(x) = P\{X \leqslant x\}$$

其中 x 为任一实数. 一个取值有限或至多可数多个的随机变量称为是离散的. 对于一个离散的随机变量 X, 我们定义其概率分布函数 $p(x)$ 为

$$p(x) = P\{X = x\}$$

如果 X 是离散随机变量, 且可能取 $x_1, x_2, \cdots$ 中的一个, 由于 X 必取这些值中的一个, 则

9

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

例 2c 假设 X 取值为 1, 2, 3 中的一个，且

$$p(1)=\frac{1}{4},\ p(2)=\frac{1}{3}$$

则由 $p(1)+p(2)+p(3)=1$ 可知 $p(3)=5/12$. □

尽管一个离散随机变量可至多取可数多个值，但我们还会经常遇到一些取值在一个区间中的随机变量. 称一个随机变量 X 是连续的随机变量，如果存在一个定义于实数域的非负函数 $f(x)$，对于任一实数集合 C，有

$$P\{X\in C\}=\int_C f(x)\mathrm{d}x \tag{2.1}$$

其中 $f(x)$ 称为随机变量 X 的概率密度函数.

累积分布函数 $F(\cdot)$ 和概率密度函数 $f(\cdot)$ 间的关系为

$$F(a)=P\{X\in(-\infty,a]\}=\int_{-\infty}^{a} f(x)\mathrm{d}x$$

对上式两边微分得到

$$\frac{\mathrm{d}}{\mathrm{d}a}F(a)=f(a)$$

这就是说，密度是累积分布函数的导数. 由等式 (2.1) 可以得到如下的密度函数的直观解释：

$$P\left\{a-\frac{\epsilon}{2}\leqslant X\leqslant a+\frac{\epsilon}{2}\right\}=\int_{a-\epsilon/2}^{a+\epsilon/2} f(x)\mathrm{d}x\approx\epsilon f(a)$$

其中 ϵ 为一小数. 换句话说，随机变量 X 落在中心为 a、长度为 ϵ 的区间内的概率近似为 $\epsilon f(a)$. 由此可以看出，$f(a)$ 可用来衡量此随机变量取值于 a 附近的可能性.

在许多随机试验中，我们的兴趣不仅在于单个随机变量的概率分布函数，而且也在于两个或多个随机变量间的关系. 为了详细说明两个随机变量 X 和 Y 间的关系，我们定义二者间的联合累积概率分布函数为

10

$$F(x,y)=P\{X\leqslant x,Y\leqslant y\}$$

即 $F(x,y)$ 给出了 X 小于等于 x 且 Y 小于等于 y 的概率.

如果 X 和 Y 是离散的随机变量，则我们定义二者的联合概率分布函数为

$$p(x,y)=P\{X=x,Y=y\}$$

类似地，称 X 和 Y 的联合分布是连续的，且其概率密度函数为 $f(x,y)$，如果对于任两个实数集 C 和 D，有

$$P\{X\in C,Y\in D\}=\iint_{x\in C,y\in D} f(x,y)\mathrm{d}x\mathrm{d}y$$

称两个随机变量 X 和 Y 是独立的，如果对于任意两个实数集 C 和 D，有

$$P\{X \in C, Y \in D\} = P\{X \in C\}P\{Y \in D\}$$

也就是说，如果对于任意两个集合 C 和 D，事件 $A = \{X \in C\}$ 和 $B = \{Y \in D\}$ 是独立的，则 X 和 Y 是独立的. 不严格地讲，如果一个随机变量的取值并不影响另一个的概率分布，则二者是独立的. 不独立的随机变量称为是相关的.

对于离散随机变量 X 和 Y，我们利用概率公理可以证明：二者独立当且仅当对于所有的 x 和 y，有

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}$$

类似地，对于两个联合的连续随机变量 X 和 Y，$f(x,y)$ 为其密度函数，则二者独立当且仅当对于所有的 x 和 y，有

$$f(x, y) = f_X(x)f_Y(y)$$

其中 $f_X(x), f_Y(y)$ 分别是 X 和 Y 的密度函数.

2.5 期 望

期望是概率论中最有用的概念之一. 如果 X 是取值为 $x_1, x_2, \cdots$ 的离散随机变量，则它的期望或期望值定义为 11

$$E[X] = \sum_i x_i P\{X = x_i\} \tag{2.2}$$

也称之为 X 的均值. 换句话说，X 的期望就是其所有可能取值的加权平均，其权为取相应值的概率. 例如，如果 X 的概率分布函数为

$$p(0) = \frac{1}{2} = p(1)$$

则

$$E[X] = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$$

就是 X 取值的平均值. 另一方面，如果

$$p(0) = \frac{1}{3},\ p(1) = \frac{2}{3}$$

则

$$E[X] = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$

是其两个可能取值 0,1 的加权平均. 因为 $p(1) = 2p(0)$, 1 的权重是 0 的 2 倍.

例 2d　如果 I 是事件 A 的示性随机变量, 即

$$I = \begin{cases} 1 & \text{如果}A\text{发生} \\ 0 & \text{如果}A\text{不发生} \end{cases}$$

则

$$E[I] = 1 \times P(A) + 0 \times P(A^C) = P(A)$$

由此可知, 某事件的示性随机变量的期望就是此事件发生的概率. □

如果 X 是一个概率密度函数为 f 的连续随机变量, 则类似于等式 (2.2), 我们定义其期望为

12

$$E[X] = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

例 2e　如果 X 的概率密度为

$$f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{其他情况} \end{cases}$$

则

$$E[X] = \int_0^1 3x^3\mathrm{d}x = \frac{3}{4}$$

□

现假设我们不想确定随机变量 X 本身的期望, 而是想确定其函数 $g(X)$ 的期望, 其中 g 是一个给定的函数. 因为当 X 取值 x 时, $g(X)$ 取值 $g(x)$, 故直观上看, $E[g(X)]$ 应是其可能取值 $g(x)$ 的加权平均, X 取值 x 的概率 (或连续时的概率密度) 即为对应于 $g(x)$ 的权. 事实上, 可以证明上述直观理解是正确的, 即有下面的结论.

命题　*如果 X 是一个概率分布函数为 $p(x)$ 的离散随机变量, 则*

$$E[g(X)] = \sum_x g(x)p(x)$$

如果 X 是一个概率密度函数为 $f(x)$ 的连续随机变量, 则

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$$

由上述命题有如下结论.

推论 如果 a, b 是常数，则

$$E[aX+b] = aE[X]+b$$

证明 对于离散情形

$$\begin{aligned} E[aX+b] &= \sum_x (ax+b)p(x) \\ &= a\sum_x xp(x) + b\sum_x p(x) \\ &= aE[X]+b \end{aligned}$$

由于连续情形的证明与此类似，故结论得证. □ 13

可以证明，对于任两个随机变量 X_1 和 X_2，有

$$E[X_1+X_2] = E[X_1]+E[X_2]$$

并且可以把它推广到多个变量的情形

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

2.6 方 差

尽管随机变量 X 的期望 $E[X]$ 是其可能取值的加权平均，但对于这些值的波动，它并没有提供任何信息. X 到 $E[X]$ 的距离平方的平均值可作为此波动的一个度量. 于是我们有如下定义.

定义 如果 X 是一个均值为 μ 的随机变量，则其方差 $\mathrm{Var}(X)$ 定义为

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$

求方差的另一公式为

$$\begin{aligned} \mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= E[X^2-2\mu X+\mu^2] \\ &= E[X^2]-E[2\mu X]+E[\mu^2] \\ &= E[X^2]-2\mu E[X]+\mu^2 \\ &= E[X^2]-\mu^2 \end{aligned}$$

即

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

一个有用的等式为：对于任意常数 a, b,

14

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

其证明留作习题.

虽然随机变量和的期望等于其期望之和，但对于方差，此结论一般不成立. 然而，在一种特殊情况下此结论是对的，即当随机变量独立时. 在证明此结论之前，我们先定义两个随机变量间的协方差.

定义 两个随机变量 X, Y 间的协方差 $\text{Cov}(X, Y)$ 定义为

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

其中 $\mu_x = E[X], \mu_y = E[Y]$.

上式右端展开则得到如下的关于协方差的有用表达式

$$\begin{aligned}\text{Cov}(X, Y) &= E[XY - \mu_x Y - X\mu_y + \mu_x\mu_y] \\ &= E[XY] - \mu_x E[Y] - E[X]\mu_y + \mu_x\mu_y \\ &= E[XY] - E[X]E[Y] \end{aligned} \tag{2.3}$$

下面我们将导出用单个方差和协方差表示的 $\text{Var}(X + Y)$ 的表达式. 因为

$$E[X + Y] = E[X] + E[Y] = \mu_x + \mu_y$$

故

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y - \mu_x - \mu_y)^2] \\ &= E[(X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)] \\ &= E[(X - \mu_x)^2] + E[(Y - \mu_y)^2] + 2E[(X - \mu_x)(Y - \mu_y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned} \tag{2.4}$$

在结束本节前，我们将证明：独立随机变量和的方差等于其方差的和.

命题 如果 X, Y 是独立的随机变量，则

$$\text{Cov}(X, Y) = 0$$

且由等式 (2.4) 有

15

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

证明 由等式 (2.3) 知，我们仅需证明 $E[XY]=E[X]E[Y]$. 对于离散情形，

$$\begin{aligned}E[XY]&=\sum_j\sum_i x_iy_jP\{X=x_i,Y=y_j\}\\&=\sum_j\sum_i x_iy_jP\{X=x_i\}P\{Y=y_j\}\quad \text{由独立性}\\&=\sum_j y_jP\{Y=y_j\}\sum_i x_iP\{X=x_i\}\\&=E[Y]E[X]\end{aligned}$$

对于连续情形，可类似证明，故命题得证. □

定义两随机变量 X,Y 之间的*相关系数* $\mathrm{Corr}(X,Y)$ 为

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

2.7 切比雪夫不等式和大数定律

我们先讲马尔可夫不等式.

命题 马尔可夫不等式 *如果 X 仅取非负值，则对于任意的 $a>0$,*

$$P\{X\geqslant a\}\leqslant\frac{E[X]}{a}$$

证明 定义随机变量 Y 为

$$Y=\begin{cases}a & \text{如果}X\geqslant a\\0 & \text{如果}X<a\end{cases}$$

由于 $X\geqslant 0$, 则易得

$$X\geqslant Y$$

上式两边取期望得

$$E[X]\geqslant E[Y]=aP\{X\geqslant a\}$$

由此结论得证. □ 16

作为上述不等式推论的切比雪夫不等式指出，一个随机变量远离其均值 k 倍标准差的概率不超过 $1/k^2$, 其中标准差定义为其方差的平方根.

推论 切比雪夫不等式 *设 X 为一均值为 μ、方差为 σ^2 的随机变量，则对于任意 $k>0$, 有*

$$P\{|X-\mu|\geqslant k\sigma\}\leqslant\frac{1}{k^2}$$

证明　因为 $(X-\mu)^2/\sigma^2$ 是一非负且均值为

$$E\left[\frac{(X-\mu)^2}{\sigma^2}\right]=\frac{E[(X-\mu)^2]}{\sigma^2}=1$$

的随机变量，故由马尔可夫不等式知

$$P\left\{\frac{(X-\mu)^2}{\sigma^2}\geqslant k^2\right\}\leqslant\frac{1}{k^2}$$

由于 $(X-\mu)^2/\sigma^2\geqslant k^2$ 与 $|X-\mu|\geqslant k\sigma$ 是等价的，故结论得证. □

弱大数定律指出，一独立同分布随机变量列前 n 项的平均值与其均值的距离大于给定正值 ϵ 的概率，当 n 趋于无穷大时趋于 0. 我们现用切比雪夫不等式证明之.

定理　弱大数定律　设 $X_1, X_2, \cdots$ 是一列均值为 μ 的独立同分布随机变量列，则对于任意的 $\epsilon>0$,

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|>\epsilon\right\}\to 0 \text{ 当} n\to\infty$$

证明　本证明是在假设 $\mathrm{Var}X_i=\sigma^2<\infty$ 的条件下完成的. 注意到

$$E\left[\frac{X_1+\cdots+X_n}{n}\right]=\frac{1}{n}(E[X_1]+\cdots+E[X_n])=\mu$$

17

$$\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right)=\frac{1}{n^2}[\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)]=\frac{\sigma^2}{n}$$

其中最后一个式子用到了独立随机变量和的方差等于各随机变量的方差和这一结论.

于是，由切比雪夫不等式，对于任意正数 k，有

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geqslant\frac{k\sigma}{\sqrt{n}}\right\}\leqslant\frac{1}{k^2}$$

对于任意 $\varepsilon>0$，令 k 满足 $k\sigma/\sqrt{n}=\varepsilon$. 即 $k^2=n\varepsilon^2/\sigma^2$，得

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geqslant\epsilon\right\}\leqslant\frac{\sigma^2}{n\epsilon^2}$$

由此，定理得证. □

强大数定律是弱大数定律的推广. 强大数定律指出，以概率 1 成立

$$\lim_{n\to\infty}\frac{X_1+\cdots+X_n}{n}=\mu$$

即可以肯定地说，一个独立同分布随机变量列的均值收敛至其期望.

2.8 某些离散随机变量

在实际应用中有几种常用的随机变量，本节介绍某些离散型的.

1. 二项随机变量

假设独立地重复成功概率为 p 的试验 n 次，如以 X 记 n 次试验中的成功次数，则称 X 是一个参数为 (n,p) 的二项随机变量，其概率分布函数为

$$P_i \equiv P\{X=i\} = \begin{pmatrix} n \\ i \end{pmatrix} p^i(1-p)^{n-i}, \ i=0,1,\cdots,n, \tag{2.5}$$

其中

$$\begin{pmatrix} n \\ i \end{pmatrix} = \frac{n!}{i!(n-i)!}$$

是二项系数，它等于从 n 个元素的集合中取 i 个元素构成的不同子集的个数. 18

我们注意到，在 n 重独立试验中，在指定位置出现 i 次成功、$n-i$ 次失败的概率为 $p^i(1-p)^{n-i}$，而这些位置的选择共有 $\begin{pmatrix} n \\ i \end{pmatrix}$ 种，于是，由此验证了等式 (2.5) 的正确性.

一个参数为 $(1,p)$ 的二项随机变量 X 称为伯努利随机变量.

由于在 n 重成功概率为 p 的独立试验中，参数为 (n,p) 的二项随机变量 X 描述的是其成功的次数，于是我们可以把 X 写成

$$X = \sum_{i=1}^{n} X_i \tag{2.6}$$

其中

$$X_i = \begin{cases} 1 & \text{如果第 } i \text{ 次试验成功} \\ 0 & \text{否则} \end{cases}$$

应用事实 $X_i^2 = X_i$(因为 $0^2=0, 1^2=1$)，我们有

$$E[X_i] = P\{X_i=1\} = p$$
$$\mathrm{Var}(X_i) = E[X_i^2] - E([X_i])^2 = p - p^2 = p(1-p)$$

于是，由等式 (2.6) 得到

$$E[X]=\sum_{i=1}^{n} E[X_i]=np$$
$$\mathrm{Var}(X)=\sum_{i=1}^{n}\mathrm{Var}(X_i) \quad \text{由独立性}$$
$$=np(1-p)$$

当计算二项概率时，下述用 p_i 表示 p_{i+1} 的递推公式是非常有用的：

$$p_{i+1}=\frac{n!}{(n-i-1)!(i+1)!}p^{i+1}(1-p)^{n-i-1}$$
$$=\frac{n!(n-i)}{(n-i)!i!(i+1)}p^i(1-p)^{n-i}\frac{p}{1-p}$$
$$=\frac{n-i}{i+1}\frac{p}{1-p}p_i$$

19

2. 泊松随机变量

一个取值为 $0,1,2,\cdots$ 的随机变量 X 称为参数 $\lambda(\lambda>0)$ 的泊松分布，如果其概率分布函数由下式给出：

$$p_i=P\{X=i\}=\mathrm{e}^{-\lambda}\frac{\lambda^i}{i!},\quad i=0,1,2,\cdots$$

其中符号 $\mathrm{e}=\lim_{n\to\infty}(1+1/n)^n$ 是一个近似等于 2.7183 的数学常数.

泊松分布有着广泛的应用，其原因是，当一次试验的成功概率很小时，大量(独立或弱相依的)这样试验的成功次数近似一个泊松随机变量. 为证明这一点，设 X 是参数 (n,p) 的二项随机变量，即它是 n 重独立的成功概率为 p 的试验中成功的次数，记 $\lambda=np$，则

$$P\{X=i\}=\frac{n!}{(n-1)!i!}p^i(1-p)^{n-i}$$
$$=\frac{n!}{(n-1)!i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$$
$$=\frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}$$

由于当 n 很大且 p 很小时，有

$$\left(1-\frac{\lambda}{n}\right)^n\approx \mathrm{e}^{-\lambda},\ \frac{n(n-1)\cdots(n-i+1)}{n^i}\approx 1,\ \left(1-\frac{\lambda}{n}\right)^i\approx 1$$

故

$$P\{X=i\}\approx e^{-\lambda}\frac{\lambda^i}{i!}$$

由于二项随机变量 Y 的均值与方差为

$$E[Y]=np,\quad \text{Var}(Y)=np(1-p)\approx np,\ \text{当}\ p\ \text{很小时}$$

而对于参数为 λ 的泊松随机变量 X，我们有

$$E[X]=\text{Var}(X)=\lambda$$

由此，可直观地看出二项随机变量与泊松随机变量间的关系. 其严格证明留作习题. 20

我们可用下面的递推公式来计算泊松概率：

$$\frac{p_{i+1}}{p_i}=\frac{\frac{e^{-\lambda}\lambda^{i+1}}{(i+1)!}}{\frac{e^{-\lambda}\lambda^i}{i!}}=\frac{\lambda}{i+1}$$

或等价地有

$$p_{i+1}=\frac{\lambda}{i+1}p_i,\ i\geqslant 0$$

假设某事件发生的次数 N 是一均值为 λ 的泊松随机变量，且进一步假设此事件的发生，意味着类型 1 事件发生或类型 2 事件发生，且二者独立，相应的概率为 p 和 $1-p$. 如以 N_i 表示类型 $i(i=1,2)$ 事件发生的次数，则 $N=N_1+N_2$. 一个有用的结论为：随机变量 N_1,N_2 是独立的泊松随机变量，均值分别为

$$E[N_1]=\lambda p,\quad E[N_2]=\lambda(1-p)$$

为证明这一结论，以 n,m 表示非负整数，且考虑联合概率 $P\{N_1=n,N_2=m\}$. 由于 $P\{N_1=n,N_2=m|N\neq n+m\}=0$，于是

$$\begin{aligned}P\{N_1=n,N_2=m\}&=P\{N_1=n,N_2=m|N=n+m\}P\{N=n+m\}\\&=P\{N_1=n,N_2=m|N=n+m\}e^{-\lambda}\frac{\lambda^{n+m}}{(n+m)!}\end{aligned}$$

然而，由于在给定 $N=n+m$ 的条件下，$n+m$ 个事件中的任一个均独立地以概率 p 为类型 1 或以概率 $1-p$ 为类型 2，于是，类型 1 的数量是参数为 $(n+m,p)$ 的二项随机变量. 这样

$$\begin{aligned}P\{N_1=n,N_2=m\}&=\binom{n+m}{n}p^n(1-p)^m e^{-\lambda}\frac{\lambda^{n+m}}{(n+m)!}\\&=\frac{(n+m)!}{n!m!}p^n(1-p)^m e^{-\lambda p}e^{-\lambda(1-p)}\frac{\lambda^n\lambda^m}{(n+m)!}\\&=e^{-\lambda p}\frac{(\lambda p)^n}{n!}e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}\end{aligned}$$

21

关于 m 求和我们得到

$$P\{N_1=n\}=\sum_m P\{N_1=n,N_2=m\}$$

$$=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^n}{n!}\sum_m \mathrm{e}^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}$$

$$=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^n}{n!}$$

类似地，我们有

$$P\{N_2=m\}=\mathrm{e}^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}$$

由此即验证了 N_1,N_2 的确是独立的均值分别为 $\lambda p,\lambda(1-p)$ 的泊松随机变量.

容易将上述结论推广到包含 r 个独立类型的情况. 如以 p_i 记类型 i 事件发生的概率 ($i=1,\cdots,r$, 且 $\sum_{i=1}^{r}p_i=1$), N_i 表示类型 i 事件发生的次数, 则类似地可以证明: $N_1,\cdots,N_r$ 是独立的均值分别为

$$E[N_i]=\lambda p_i,\ i=1,\cdots,r$$

的泊松随机变量.

3. 几何随机变量

在多重独立重复且成功概率为 p 的试验列中, 以 X 记首次成功出现时的试验次数. 我们注意到, $\{X=n\}$ 表示第一次成功出现在第 n 次试验时, 即前 $n-1$ 次试验均失败, 而第 n 次试验成功, 由于这些试验是独立的, 有

$$P\{X=n\}=p(1-p)^{n-1},\ n\geqslant 1 \tag{2.7}$$

一个概率分布函数为等式 (2.7) 的随机变量称为参数 p 的几何随机变量, 其均值为

22

$$E[X]=\sum_{n=1}^{\infty}np(1-p)^{n-1}=\frac{1}{p}$$

在上面的等式中, 我们用到了下面的代数等式: 对于 $0<x<1$,

$$\sum_{n=1}^{\infty}nx^{n-1}=\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{n=0}^{\infty}x^n\right)=\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right)=\frac{1}{(1-x)^2}$$

另不难证明

$$\mathrm{Var}(X)=\frac{1-p}{p^2}$$

4. 负二项随机变量

在多重独立重复且成功概率为 p 的试验列中，如以 X 记 r 次成功出现时的试验次数，则称之为参数 (p,r) 的负二项分布或帕斯卡分布，其概率分布函数为

$$P\{X=n\}=\binom{n-1}{r-1}p^r(1-p)^{n-r},\ n\geqslant r \tag{2.8}$$

为验证等式 (2.8) 的正确性，我们注意到在前 $n-1$ 次试验中有 $r-1$ 次成功的概率为

$$\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$$

而第 n 次试验成功的概率为 p.

若以 $X_i,\ i=1,2,\cdots,r$ 记 $i-1$ 次成功后直到出现第 i 次成功需要的试验次数，则它们是独立的参数 p 的几何随机变量，因为

$$X=\sum_{i=1}^{r}X_i$$

由几何随机变量的结论可得

$$E[X]=\sum_{i=1}^{r}E[X_i]=\frac{r}{p}$$

$$\mathrm{Var}(X)=\sum_{i=1}^{r}\mathrm{Var}(X_i)=\frac{r(1-p)}{p^2}$$

23

5. 超几何随机变量

设一个盒子里有 $N+M$ 个球，其中 N 个是浅色的，M 个是深色的. 如果从盒子里随机地取 n 个，即包含 n 个球的 $\binom{N+M}{n}$ 个子集被抽中的概率相等，则浅色球个数 X 的概率分布函数为

$$P\{X=i\}=\frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

称具有上述概率分布函数的随机变量 X 为超几何随机变量.

假设这 n 个球被相继取出，如果记

$$X_i = \begin{cases} 1 & \text{如果第 } i \text{ 个球为浅色的} \\ 0 & \text{否则} \end{cases}$$

则

$$X = \sum_{i=1}^{n} X_i \tag{2.9}$$

由此可得

$$E[X] = \sum_{i=1}^{n} E[X_i] = \frac{nN}{N+M}$$

在上式的证明中利用了如下事实：由对称性可知，第 i 次取得浅球的概率与从 $N+M$ 个球中任取一球为浅色球的概率是相等的，即 $E[X_i] = P\{X_i = 1\} = \frac{N}{N+M}$.

因为 X_i 不是独立的（为什么？），则在利用等式 (2.9) 计算 $\mathrm{Var}(X)$ 时应有协方差项. 于是，由此可以求得

$$\mathrm{Var}(X) = \frac{nNM}{(N+M)^2}\left(1 - \frac{n-1}{N+M-1}\right)$$

2.9 连续随机变量

24

本节我们考虑几种连续型随机变量.

1. 均匀分布随机变量

称随机变量 X 在区间 $(a,b)(a<b)$ 上均匀分布，如果其概率密度函数为

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{如果 } a < x < b \\ 0 & \text{其他情况} \end{cases}$$

换句话说，如果一个随机变量 X 取值在区间 (a,b) 内，且取其中的任一值的概率都是相等的，则 X 在 (a,b) 上均匀分布.

(a,b) 上均匀分布随机变量 X 的期望和方差分别为

$$E[X] = \frac{1}{b-a}\int_a^b x\mathrm{d}x = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

$$E[X^2] = \frac{1}{b-a}\int_a^b x^2\mathrm{d}x = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+b^2+ab}{3}$$

$$\mathrm{Var}(X) = \frac{1}{3}(a^2+b^2+ab) - \frac{1}{4}(a^2+b^2+2ab) = \frac{1}{12}(b-a)^2$$

这样，正如人们所想像的，均匀分布的期望就是区间 (a,b) 的中点.

X 的分布函数为：对任意的 $a<x<b$，有

$$F(x)=P\{X\leqslant x\}=\int_a^x (b-a)^{-1}\mathrm{d}x=\frac{x-a}{b-a}$$

25

2. 正态随机变量

称随机变量 X 服从参数为 (μ,σ^2) 的正态分布，如果其概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2},\ -\infty<x<\infty$$

正态密度函数是一个钟形曲线且关于 μ 对称 (见图 2-1).

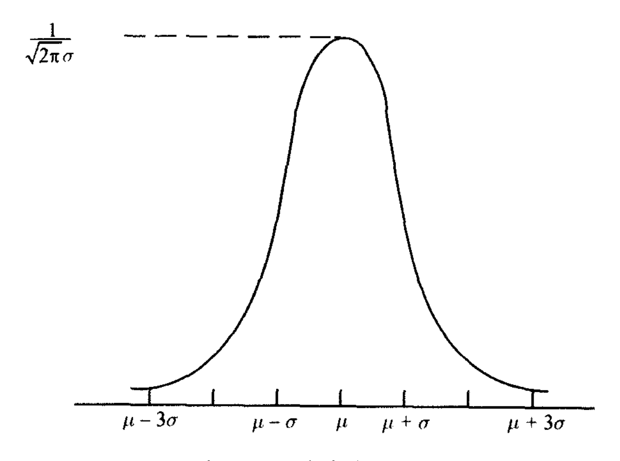


图 2-1 正态密度函数

容易证明，参数 μ,σ^2 分别是正态分布的期望与方差，即

$$E[X]=\mu \quad \mathrm{Var}(X)=\sigma^2$$

关于正态随机变量有如下一个重要的结论：如果 X 是一个均值为 μ 和方差为 σ^2 的正态随机变量，则对于任两个常数 a 和 b，$aX+b$ 是均值为 $a\mu+b$ 和方差为 $a^2\sigma^2$ 的正态随机变量. 由此结论可得，如果 X 是均值为 μ 和方差为 σ^2 的正态随机变量，则

$$Z=\frac{X-\mu}{\sigma}$$

是均值为 0 和方差为 1 的正态随机变量. 我们称这样的随机变量 Z 是标准正态的. 以 Φ 记标准正态随机变量的分布函数，即

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-x^2/2}\mathrm{d}x,\ -\infty<x<\infty$$

上述 Z 与 X 间的关系是非常有用的，因为关于 X 的所有概率都可以用 Φ 来计算. 例如，X 的分布函数可以表示为

$$
\begin{aligned}
F(x) &= P\{X \leqslant x\} \\
&= P\left\{\frac{X-\mu}{\sigma} \leqslant \frac{x-\mu}{\sigma}\right\} \\
&= P\left\{Z \leqslant \frac{x-\mu}{\sigma}\right\} \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

26 $\Phi(x)$ 的值可以查表求得，也可以编程近似求取.

对区间 $(0,1)$ 内的任一值 a，设 z_a 满足

$$P\{Z > z_a\} = 1 - \Phi(z_a) = a$$

即一个标准正态随机变量超过 z_a 的概率为 a(见图 2-2). 值 z_a 可由 Φ 的数表查得. 例如，由于

$$\Phi(1.64) = 0.95,\ \Phi(1.96) = 0.975,\ \Phi(2.33) = 0.99$$

则

$$z_{0.05} = 1.64,\ z_{0.025} = 1.96,\ z_{0.01} = 2.33$$

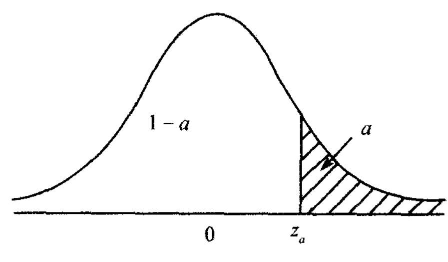


图 2-2　$P\{Z > z_a\} = a$

正态随机变量之所以有着广泛的应用，其最主要的一个原因是来自于概率论中的中心极限定理. 中心极限定理指出，大量独立同分布随机变量的和渐近正态分布，下面定理即为最简单的一个.

中心极限定理　设 $X_1, X_2, \cdots$ 是一列独立同分布的随机变量，其均值为 μ，方差为 σ^2. 则

$$\lim_{n\to\infty} P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} < x\right\} = \Phi(x)$$

3. 指数随机变量

一个概率密度函数为

$$f(x) = \lambda \mathrm{e}^{-\lambda x},\ 0 < x < \infty$$

的随机变量 X 称作参数为 $\lambda(\lambda > 0)$ 的指数随机变量，其累积分布为 [27]

$$F(x) = \int_0^x \lambda \mathrm{e}^{-\lambda x} \mathrm{d}x = 1 - \mathrm{e}^{-\lambda x},\ 0 < x < \infty$$

易证其期望与方差分别为

$$E[X] = \frac{1}{\lambda} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

无记忆性是指数随机变量的关键特性. 称一个非负随机变量 X 是无记忆的，如果

$$P\{X > s + t | X > s\} = P\{X > t\}\ \forall\ s, t \geqslant 0 \tag{2.10}$$

之所以称上述性质为无记忆性，我们可以如下理解：以 X 表示某产品的寿命，现考虑已知其寿命大于 s 的条件下，它能再活 t 个时间单位的概率，即求在时刻 s 它“活着”的条件下，其寿命超过 $t+s$ 的概率，于是

$$P\{一个年龄为s的产品能再活t个时间单位\} = P\{X > s + t | X > s\}$$

这样，等式 (2.10) 指出，年龄为 s 的产品的寿命与 s 无关，也就是说，为求此产品的剩余寿命，我们无需记住其现在的年龄.

等式 (2.10) 等价于

$$P\{X > s + t\} = P\{X > s\} P\{X > t\}$$

只要 X 是一个指数随机变量，则由 $P\{X > x\} = \mathrm{e}^{-\lambda x}$ 可知它满足上式，即指数随机变量是无记忆的.（事实上，不难证明指数随机变量是惟一无记忆的连续随机变量.）

指数随机变量的另一有用性质是其与一个正的常数的乘积仍是一个指数随机变量. 为证明这一点，设 X 是参数 λ 的指数随机变量，c 为一正数，则

$$P\{cX \leqslant x\} = P\{X \leqslant x/c\} = 1 - \mathrm{e}^{-\lambda x/c}$$

即说明 cX 是参数为 λ/c 的指数随机变量.

设 $X_1, \cdots, X_n$ 是独立的参数分别为 $\lambda_1, \cdots, \lambda_n$ 的指数随机变量，则一个有用的结果为：$\min(X_1, \cdots, X_n)$ 是参数为 $\sum_i \lambda_i$ 的指数随机变量，且与 X_i 中最小的那 [28]

个随机变量独立. 事实上, 记 $M=\min(X_1,\cdots,X_n)$, 则

$$
\begin{aligned}
P\{X_j=\min_i X_i|M>t\}&=P\{X_j-t=\min_i(X_i-t)|M>t\}\\
&=P\{X_j-t=\min_i(X_i-t)|X_i>t,i=1,\cdots,n\}\\
&=P\{X_j=\min_i X_i\},
\end{aligned}
$$

上式最后一个不等式是由指数随机变量的无记忆性得到的. 我们注意到, 在已知 X_i 大于 t 的条件下, X_i-t 仍是参数 λ_i 的指数随机变量. 于是, 在已知所有的 X_i 大于 t 的条件下, $X_1-t,\cdots,X_n-t$ 的条件分布与 $X_1,\cdots,X_n$ 的无条件分布相同. 由此可知, M 与 X_i 中最小者独立.

因为

$$
P\{M>t\}=P\{X_i>t,\ i=1,\cdots,n\}=\prod_{i=1}^{n}P\{X_i>t\}=\mathrm{e}^{-\sum_{i=1}^{n}\lambda_i t}
$$

故 M 的分布是参数为 $\sum_i\lambda_i$ 的指数分布.

X_i 是最小值的概率为

$$
\begin{aligned}
P\{X_j=M\}&=\int P\{X_j=M|X_j=t\}\lambda_j\mathrm{e}^{-\lambda_j t}\mathrm{d}t\\
&=\int P\{X_i>t,i\neq j|X_j=t\}\lambda_j\mathrm{e}^{-\lambda_j t}\mathrm{d}t\\
&=\int P\{X_i>t,i\neq j\}\lambda_j\mathrm{e}^{-\lambda_j t}\mathrm{d}t\\
&=\int\left(\prod_{i\neq j}\mathrm{e}^{-\lambda_i t}\right)\lambda_j\mathrm{e}^{-\lambda_j t}\mathrm{d}t\\
&=\lambda_j\int\mathrm{e}^{-\sum\limits_i\lambda_i t}\mathrm{d}t=\frac{\lambda_j}{\sum\limits_i\lambda_i}
\end{aligned}
$$

4. 泊松过程和伽玛随机变量

假设某事件流在随机时间点上发生, 且以 $N(t)$ 记此事件流在时间 $[0,t]$ 内发生的次数. 称此事件流构成一个参数 $\lambda(\lambda>0)$ 的泊松过程, 如果

(a) $N(0)=0$.

[29] (b) 在不相交时间段内发生的事件流的个数独立.

(c) 在给定区间内发生的事件流个数的分布仅依赖于此区间的长度而与位置无关.

(d) $\lim_{h\to 0} \frac{P\{N(h)=1\}}{h} = \lambda$.

(e) $\lim_{h\to 0} \frac{P\{N(h)\geqslant 2\}}{h} = 0$.

条件 (a) 说明此过程从时间 0 开始. 独立增量假设 (b) 指出, 时刻 t 的事件流数 $N(t)$ 与时刻 t 和 $t+s$ 间的事件流数 $N(t+s)-N(t)$ 独立. 平稳增量假设 (c) 指出 $N(t+s)-N(t)$ 的分布与 t 无关. 条件 (d) 和 (e) 指出在长度为 h 的小区间内, 发生一个事件的概率近似为 λh, 而发生两个或更多事件的概率近似为 0.

由这些条件可以证明, 发生在长度为 t 的时间区间内的事件数是一个参数为 λt 的泊松随机变量. 为证明这一点, 考虑区间 $[0,t]$, 并把它划分成 n 个长度分别为 t/n 的不相交子区间 (见图 2-3). 首先考虑包含一个事件的子区间个数. 因为由条件 (b) 和 (c) 可知, 这些不相交子区间相互独立地以相同的概率 p 包含一个事件, 且 $p\approx \lambda t/n$, 所以, 这些子区间的个数是一个参数为 (n,p) 的二项随机变量. 于是, 由二项分布收敛于泊松分布这一结论可知, 当 $n\to\infty$ 时, 包含一个事件的子区间个数收敛于参数为 λt 的泊松随机变量. 又由条件 (e) 可知, 当 $n\to\infty$ 时, 任一子区间包含两个或更多事件的概率趋于 0. 综合以上两点, 在区间 $[0,t]$ 内出现的事件数 $N(t)$ 是一参数为 λt 的泊松随机变量.

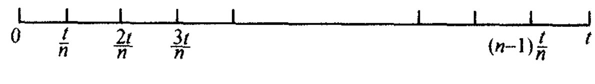


图 2-3 区间 $[0,t]$ 的划分

以 X_1 记一个泊松过程中首个事件发生的时间, 以 X_n 记从第 $n-1$ 个事件发生到第 n 个事件发生的时间间隔 $(n>1)$, 则称 $\{X_n,\ n=1,2,\cdots\}$ 为到达时隔序列. 例如, 如果 $X_1=5, X_2=10$, 则此泊松过程的第一个事件将在时刻 5 出现, 第二个事件将在时刻 15 出现.

现确定 X_n 的分布. 为此, 我们先注意到当且仅当在区间 $[0,t]$ 内无此泊松过程的事件出现时, 事件 $\{X_1>t\}$ 发生, 即有

$$P\{X_1>t\}=P\{N(t)=0\}=\mathrm{e}^{-\lambda t}$$

于是, X_1 服从均值为 $1/\lambda$ 的指数分布. 为得到 X_2 的分布, 注意到

$$\begin{aligned}P\{X_2>t|X_1=s\}&=P\{\text{在}(s,s+t]\text{内无事件出现}|X_1=s\}\\&=P\{\text{在}(s,s+t]\text{内无事件出现}\}\\&=\mathrm{e}^{-\lambda t}\end{aligned}$$

其中后两个等式是由独立增量和平稳增量得到的. 由上式可知, X_2 也服从均值为 $1/\lambda$ 的指数分布且与 X_1 独立. 重复上面内容, 则有如下结论:

命题 到达时间间隔 $X_1, X_2, \cdots$ 是独立同分布的参数为 λ 的指数随机变量列.

以 $S_n = \sum_{i=1}^{n} X_i$ 记第 n 个事件出现的时刻. 因为 $S_n \leqslant t$ 当且仅当时刻 t 时至少有 n 个事件已经出现, 所以,

$$P\{S_n \leqslant t\} = P\{N(t) \geqslant n\} = \sum_{j=n}^{\infty} \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!}$$

由于上式左端是 S_n 的累积分布函数, 故两端求导后得到 S_n 的密度函数 $f_n(t)$ 如下:

$$\begin{aligned} f_n(t) &= \sum_{j=n}^{\infty} j\lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{j-1}}{j!} - \sum_{j=n}^{\infty} \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \sum_{j=n}^{\infty} \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

定义 称具有概率密度函数

$$f(t) = \lambda \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \ t > 0$$

的随机变量是参数为 (n, λ) 的伽玛随机变量.

综上所述, 对于一个参数为 λ 的泊松过程, 其第 n 个事件发生的时间 S_n 是一个参数为 (n, λ) 的伽玛随机变量. 又因 $S_n = \sum_{i=1}^{n} X_i$, 故由上一命题有如下推论.

31

推论 n 个独立的、参数为 λ 的指数随机变量之和是一个参数为 (n, λ) 的伽玛随机变量.

5. 非齐次泊松过程

从模拟的角度看, 上述泊松过程最主要的不足在于它假设事件等可能地出现在任何长度相等的区间内. 如放松这一条件, 则导致非齐次或非平稳的泊松过程.

设事件的发生是随机的, 且以 $N(t)$ 记时刻 t 时出现的事件数, 则我们称 $\{N(t), t \geqslant 0\}$ 构成一个强度函数为 $\lambda(t), t \geqslant 0$ 的非齐次泊松过程, 如果

(a) $N(0) = 0$.

(b) 发生在不相交时间区间的事件个数是独立的.

(c) $\lim\limits_{h \to 0} P\{$ 仅一个事件出现在区间 $(t, t+h)$ 内$\}/h = \lambda(t)$.

(d) $\lim\limits_{h \to 0} P\{$ 两个或更多事件出现在区间 $(t, t+h)$ 内$\}/h = 0$.

称函数

$$m(t) = \int_0^t \lambda(s)\mathrm{d}s,\ t \geqslant 0$$

为均值函数. 我们有如下命题.

命题 $N(t+s) - N(t)$ 是参数为 $m(t+s) - m(t)$ 的泊松随机变量.

被称为时刻 t 的强度 $\lambda(t)$ 反映了一个事件在时刻 t 发生的可能性. 我们注意到, 如果 $\lambda(t) \equiv \lambda$, 则非齐次泊松过程就是通常的泊松过程. 下一命题给出了非齐次泊松过程的一个很有用的解释.

命题 假设事件流服从参数为 λ 的泊松过程, 且无论时刻 t 前发生了什么, 我们均假设一个在时刻 t 发生的事件以概率 $p(t)$ 被记录. 则被记录到的事件流就构成一个强度函数为 $\lambda(t) = \lambda p(t)$ 的非齐次泊松过程.

证明 我们注意到, 本命题满足非齐次泊松过程所要求的条件 (a), (b) 和 (d), 而条件 (c) 可由下式得到: 32

$$\begin{aligned}
&P\{\text{在区间}(t,t+h)\text{内, 一个事件被记录到}\}\\
=&P\{\text{发生一个事件并且它被记录到}\}\\
&+P\{\text{发生两个或更多个事件, 而仅一个事件被记录到}\}\\
\approx&\lambda h p(t)
\end{aligned}$$

□

2.10 条件期望与条件方差

设 X, Y 是两离散型随机变量, 我们定义给定 $Y=y$ 下 X 的条件期望 $E[X|Y=y]$ 如下:

$$\begin{aligned}
E[X|Y=y] &= \sum_x xP\{X=x|Y=y\}\\
&= \frac{\sum_x xP\{X=x, Y=y\}}{P\{Y=y\}}
\end{aligned}$$

换言之, 给定 $Y=y$ 下 X 的条件期望仍和以前一样, 它是 X 取值的加权平均, 只是此时值 x 的权取为 Y 等于 y 下 X 等于 x 的条件概率.

类似地, 如果 X, Y 是密度函数为 $f(x,y)$ 的连续型随机变量, 则给定 $Y=y$ 下 X 的条件期望定义为

$$E[X|Y=y] = \frac{\int xf(x,y)\mathrm{d}x}{\int f(x,y)\mathrm{d}x}$$

以 $E[X|Y]$ 记随机变量 Y 的函数，其在 $Y=y$ 点的取值为 $E[X|Y=y]$. 注意到 $E[X|Y]$ 是一随机变量，且有如下非常有用的结论.

命题

$$E[E[X|Y]]=E[X] \tag{2.11}$$

如果 Y 是离散的随机变量，则由等式 (2.11) 知

33

$$E[X]=\sum_y E[X|Y=y]P\{Y=y\}$$

如果 Y 是连续的随机变量，g 为其密度，则等式 (2.11) 指出

$$E[X]=\int E[X|Y=y]g(y)\mathrm{d}y$$

下面我们给出 X,Y 均为离散时的证明.

$$\begin{aligned}\sum_y E[X|Y=y]P\{Y=y\}&=\sum_y\sum_x xP\{X=x|Y=y\}P\{Y=y\}\\&=\sum_y\sum_x xP\{X=x,Y=y\}\\&=\sum_x x\sum_y P\{X=x,Y=y\}\\&=\sum_x xP\{X=x\}=E[X]\end{aligned}$$

给定 Y 下 X 的条件方差定义为

$$\mathrm{Var}(X|Y)=E[(X-E[X|Y])^2|Y]$$

即 $\mathrm{Var}(X|Y)$ 是 Y 的函数，其在 $Y=y$ 点的取值为 $Y=y$ 下 X 的方差. 用得到 $\mathrm{Var}(X)=E[X^2]-(E[X])^2$ 同样的方法，我们有

$$\mathrm{Var}(X|Y)=E[X^2|Y]-(E[X|Y])^2$$

在上式两边取期望，有

$$\begin{aligned}E[\mathrm{Var}(X|Y)]&=E[E[X^2|Y]]-E[(E[X|Y])^2]\\&=E[X^2]-E[(E[X|Y])^2]\end{aligned} \tag{2.12}$$

又因为 $E[E[X|Y]]=E[X]$，故我们有

$$\mathrm{Var}(E[X|Y])=E[(E[X|Y])^2]-(E[X])^2 \tag{2.13}$$

结合等式 (2.12) 和 (2.13)，我们得出如下被称为条件方差公式的等式.

条件方差公式

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

34

习 题

1. (a) 对任意事件 A, B，证明

$$A \cup B = A \cup A^{\mathrm{c}}B, \quad B = AB \cup A^{\mathrm{c}}B$$

(b) 证明

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

2. 考虑编号为 1 至 6 的六匹马的赛跑试验，假设样本空间为

$$S = \{(1,2,3,4,5,6)\text{的所有排列}\}$$

记 A 为 1 号赛马取得前三名的事件，B 为 2 号赛马取得第二名的事件，C 为 3 号赛马取得第三名的事件.

(a) 描述 $A \cup B$. 此事件包含多少个试验结果？

(b) 事件 AB 包含多少个试验结果？

(c) 事件 ABC 包含多少个试验结果？

(d) 事件 $A \cup BC$ 包含多少个试验结果？

3. 一对夫妻有两个孩子，假设其四种可能的概率相等. 当已知大的为女孩时，两个都是女孩的概率是多少？

4. 已知国王来自有两个孩子的家庭，请问此家庭另一个孩子是国王兄弟的概率.

5. 随机变量 X 取值 $1,2,3,4$，其概率为

$$P\{X = i\} = ic, \quad i = 1,2,3,4$$

其中 c 为常数，求 $P\{2 \leqslant X \leqslant 3\}$.

6. 连续随机变量 X 的概率密度函数为

$$f(x) = cx, \quad 0 < x < 1$$

求 $P\{X > 1/2\}$.

7. X, Y 的联合概率密度函数为

$$f(x,y) = 2\mathrm{e}^{-(x+2y)}, \quad 0 < x < \infty,\ 0 < y < \infty$$

求 $P\{X < Y\}$.

35

8. 求由习题 5 定义的随机变量的期望.
9. 求由习题 6 定义的随机变量的期望.
10. 设有 10 种不同类型的优惠券，某人每次均以等概率取得其中一张. 以 X 记此人所得到的 N 张优惠券中不同类型数，求 $E[X]$. [提示：对于 $i=1,\cdots,10$，记

$$X_t = \begin{cases} 1 & \text{如果第 } i \text{ 类优惠券在} N \text{张之中} \\ 0 & \text{其他情况} \end{cases}$$

利用式子 $X=\sum_{i=1}^{10} X_i$.]
11. 一个六面的色子，如果各点出现的概率相同，求其出现点数的方差.
12. 设 X 的概率密度函数为

$$f(x) = c\mathrm{e}^{x},\ 0 < x < 1$$

求 $\mathrm{Var}(X)$.
13. 证明 $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.
14. 设装在某容器内苹果酱的数量 X 是均值为 4 克的随机变量.
    (a) 我们能给出此罐有多于 6g 苹果酱的概率吗？
    (b) 如果 $\mathrm{Var}(X) = 4\mathrm{g}^2$，我们能给出此罐有 3 至 5g 苹果酱的概率吗？
15. 至少需一半发动机正常运转才能保证一架飞机的安全飞行，现假设每台发动机运转正常与否是独立的，且正常运转的概率为 p，试问当 p 取何值时具有 3 个发动机的飞机比具有 5 个发动机的飞机安全？
16. 对于参数为 (n,p) 的二项随机变量 X，证明 $P\{X=i\}$ 先增后减，其最大值在 i 为不超过 $(n+1)p$ 的最大整数时达到.
17. 设 X,Y 是两参数分别为 $(n,p),(m,p)$ 的独立的二项随机变量，不通过计算证明 $X+Y$ 是参数为 $(n+m,p)$ 的二项随机变量.

36

18. 请解释为什么下面的随机变量均近似服从泊松分布：
    (a) 本书某章印刷错误数.
    (b) 一天拨错电话号码的次数.
    (c) 某天去一指定邮局的顾客人数.
19. 设 X 是一参数为 λ 的泊松随机变量，证明
    (a) $E[X]=\lambda$.
    (b) $\mathrm{Var}(X)=\lambda$.
20. 设 X,Y 独立，且是参数分别为 λ_1,λ_2 的泊松随机变量，请利用习题 17 的结果，说明 $X+Y$ 是参数为 $\lambda_1+\lambda_2$ 的泊松随机变量，并给出分析证明.

[提示：$P\{X+Y=k\} = \sum_{i=0}^{k} P\{X=i, Y=k-i\} = \sum_{i=0}^{k} P\{X=i\}P\{Y=k-i\}$]

21. 解释如何利用关系式

$$p_{i+1}=\frac{\lambda}{i+1}p_i$$

有效计算泊松概率.

22. 设 X 是参数为 p 的几何随机变量, 求 $P\{X>n\}$.

23. 设两位选手玩某种游戏, 若有一人总共赢取五场胜利, 则游戏结束. 设选手 A 在每场游戏中取胜的概率为 0.6, 则他最终取胜的概率为多少?

24. 考虑 2.8 节中的超几何模型, 并把所有白球都编号. 对于 $i=1,\cdots,N$, 记

$$Y_i=\begin{cases}1 & \text{若抽中编号为 } i \text{ 的白球}\\ 0 & \text{其他情况}\end{cases}$$

证明 $X=\sum_{i=1}^{N}Y_i$, 并利用此表达式求 $E[X]$, 再验证此结果与 2.8 所给结果一致.

25. 设公共汽车在早八点至八点半之间到达某车站的时间为均匀随机变量, 如果早八点到达此站, 请问要等候 5 至 15 分钟的概率为多少?

26. 对于参数为 (μ,σ^2) 的正态随机变量 X, 请证明

(a) $E[X]=\mu$.

(b) $\mathrm{Var}(X)=\sigma^2$.

37

27. 设 X 是参数 (n,p) 的二项随机变量, 解释为什么对于大的 n 有

$$P\left\{\frac{X-np}{\sqrt{np(1-p)}}\leqslant x\right\}\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-x^2/2}\mathrm{d}x$$

28. 对于参数为 λ 的指数随机变量 X, 请证明

(a) $E[X]=1/\lambda$.

(b) $\mathrm{Var}(X)=1/\lambda^2$.

29. 一天早晨, A,B,C 三人在一个只有两个出纳员的银行等候服务. 假设 A,B 两人已分别到两个窗口接受服务而 C 仍在排队. 如果一位顾客所需服务时间是一参数为 λ 的指数随机变量, 则 C 为最后离开银行的人的概率是多少? [提示: 不需要计算.]

30. 设 X,Y 是两个独立的指数随机变量, 参数分别为 λ,μ, 问 $\max(X,Y)$ 仍是指数随机变量吗?

31. 对于一个事件发生速率为每小时 0.3 次的泊松过程, 在 10 点至 14 点之间没有事件发生的概率是多少?

32. 对于一个参数为 λ 的泊松过程, 计算 $P\{N(s)=k|N(t)=n\}$, 其中 $s<t$.

33. 计算当 $s>t$ 时习题 32 中的概率.

34. 对于参数为 (n,λ) 的伽玛随机变量 X, 求

(a) $E[X]$.

(b) $\mathrm{Var}(X)$.

35. 在一个盒子里有 4 个白球和 6 个黑球，从中随机地抽出 4 个球，以 X 记其白球数. 现从剩余的 6 个球中再随机抽一个，若此球为白球，则 Y 等于 1；若此球为黑球，则 Y 等于 0. 计算：
   (a) $E[Y|X=2]$.
   (b) $E[X|Y=1]$.
   (c) $\text{Var}(Y|X=0)$.
   (d) $\text{Var}(X|Y=1)$.

36. 如果 X, Y 是独立同分布的指数随机变量，证明：在给定 $X+Y=t$ 下，X 的条件分布是 $(0,t)$ 上的均匀分布.

38

参考文献 [1]

[1] Feller, W., *An Introduction to Probability Theory and Its Applications*, 3rd ed. Wiley. New York, 1968.

[2] Ross, S. M., *A First Course in Probability*, 6th ed. Prentice Hall, New Jersey, 2002.

[3] Ross, S. M., *Introduction to Probability Models*, 7th ed. Academic Press, New York, 2000.

39

1. 此三本书均由人民邮电出版社翻译出版. 详情请查询图灵公司网址 www.turingbook.com.
—— 编者注

第3章 随 机 数

引 言

随机数是指 (0,1) 上均匀分布随机变量的值，随机数的产生是随机模拟的基石. 本章我们将介绍如何利用计算机产生随机数，并举例说明其应用.

3.1 伪随机数的产生

随机数最早是手工或通过机械方式由手纺车、掷骰子或洗纸牌等产生的，而现在则由计算机相继产生伪随机数. 尽管作为数列的伪随机数是机器生成的，但它们具有 (0,1) 上均匀分布独立随机变量的一切特征.

一种最常用的产生伪随机数的方法是：从初值 x_0（也称为种子）出发，利用公式

$$x_n = ax_{n-1} \mod m \tag{3.1}$$

逐步计算 x_n, $n \geqslant 1$, 其中 a 和 m 是给定的正整数，上式表示 x_n 为 ax_{n-1} 被 m 除后的余数. 于是，每一个 x_n 均为 $0, 1, \cdots, m-1$ 中的一个，x_n/m 称为一个伪随机数，它近似服从 (0,1) 上的均匀分布.

称由等式 (3.1) 产生随机数的方法为乘同余法. 由于每一个 x_n 均取值 $0, 1, \cdots, m-1$ 中的一个，故若干次后（至多 m 次）所产生的随机数必定重复，且自此之后整个序列也开始重复. 于是，我们在选择常数 a 及 m 时，希望对于任一初值 x_0，在重复出现前所产生的随机数序列足够长.

一般地，在选取常数 m 及 a 时，应遵循如下三个准则：

1. 对于任一初值，产生的序列具有来自 $(0,1)$ 上均匀分布独立随机变量列的特征.

2. 对于任一初值，在重复出现前所产生的随机数序列足够长.

3. 每一数值均可由计算机有效计算.

为满足上述三个条件，可选一个符合计算机字长要求的较大的素数作为 m. 对于 32 位计算机 (其中第一个字节为正负号)，已经证明 $m = 2^{31}-1$ 和 $a = 7^5 = 16\,807$ 符合上述要求 (对于 36 位计算机，取 $m = 2^{35} - 31, a = 5^5$ 为好).

另一个生成伪随机数的方法是如下的递推公式：

$$x_n = (ax_{n-1} + c) \mod m$$

由于它包括加法与乘法，故此方法被称为混合同余法. 当采用这种方法生成随机数时，人们经常取 m 为计算机字长，这是由于这种取法易于非常有效地计算 $(ax_{n-1} + c) \mod m$，即 $(ax_{n-1} + c)/m$.

由于本书的出发点是假设我们可以生成一列近似于独立的 $(0,1)$ 上均匀分布随机变量列的伪随机数，故我们将不在本书中研究有关的理论问题，如"好"的伪随机数的构造问题，而假设我们有一个可以产生所求随机数的黑盒子.

3.2　利用随机数求积分

随机数的最早应用之一就是积分计算. 设 $g(x)$ 是一个函数，我们想计算如下的 θ

42

$$\theta = \int_0^1 g(x)\mathrm{d}x$$

为计算 θ，我们注意到如果 U 在 $(0,1)$ 上均匀分布，则 θ 可表示为

$$\theta = E[g(U)]$$

如果 $U_1, \cdots, U_k$ 独立且为 $(0,1)$ 上均匀分布的随机变量，则 $g(U_1), \cdots, g(U_k)$ 是独立同分布的随机变量，且其均值为 θ. 于是，由强大数定律，我们以概率 1 有

$$\sum_{i=1}^{k} \frac{g(U_i)}{k} \to E[g(U)] = \theta, \quad 当k \to \infty$$

这样，我们可以如下近似计算 θ：先产生大量的随机数 u_i，之后用 $g(u_i)$ 的平均值近似 θ. 近似计算积分的这种方法被称为蒙特卡洛方法.

如果我们想计算

$$\theta = \int_a^b g(x)\mathrm{d}x$$

则通过变量替换 $y = (x-a)/(b-a), \mathrm{d}y = \mathrm{d}x/(b-a)$，我们有

$$\theta = \int_0^1 g(a + [b-a]y)(b-a)\mathrm{d}y = \int_0^1 h(y)\mathrm{d}y$$

其中 $h(y) = (b-a)g(a + [b-a]y)$. 于是，我们可以通过求 h 在多个随机数处的平均值来近似 θ.

类似地，如果我们想计算

$$\theta = \int_0^\infty g(x)\mathrm{d}x$$

则可以通过替换 $y = 1/(x+1), \mathrm{d}y = -\mathrm{d}x/(x+1)^2 = -y^2\mathrm{d}x$ 得到等式

$$\theta = \int_0^1 h(y)\mathrm{d}y$$

其中

$$h(y) = \frac{g(\frac{1}{y} - 1)}{y^2}$$

43

显然，我们也可以把上述利用随机数求取一元积分的方法推广到多重积分. 设 g 是一 n 元函数，我们感兴趣的积分为

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 g(x_1, \cdots, x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$$

利用蒙特卡洛方法估计 θ 的关键在于 θ 可以表示成下面的期望形式

$$\theta = E[g(U_1, \cdots, U_n)]$$

其中 $U_1, \cdots, U_n$ 是独立的 (0,1) 上均匀分布的随机变量. 于是，如果我们先产生 k 组独立的集合，其中每组包含 n 个独立的 (0,1) 上均匀分布的随机变量

$$\begin{gathered} U_1^1, \cdots, U_n^1 \\ U_1^2, \cdots, U_n^2 \\ \vdots \\ U_1^k, \cdots, U_n^k \end{gathered}$$

由于 $g(U_1^i, \cdots, U_n^i)$ $(i = 1, \cdots, k)$ 是独立同分布的随机变量且其均值为 θ，故我们可以用 $\sum_{i=1}^k g(U_1^i, \cdots, U_n^i)/k$ 来估计 θ.

下面考虑利用上述方法估计 π.

例 3a　π 的估计　假设随机向量 (X, Y) 在中心为原点、面积为 4 的正方形上均匀分布，即它是图 3-1 所给区域内的一个随机点. 现考虑此正方形上的随机点落在半径为 1 的所给圆内的概率（见图 3-2）.

我们注意到 (X, Y) 在此正方形内均匀分布，故

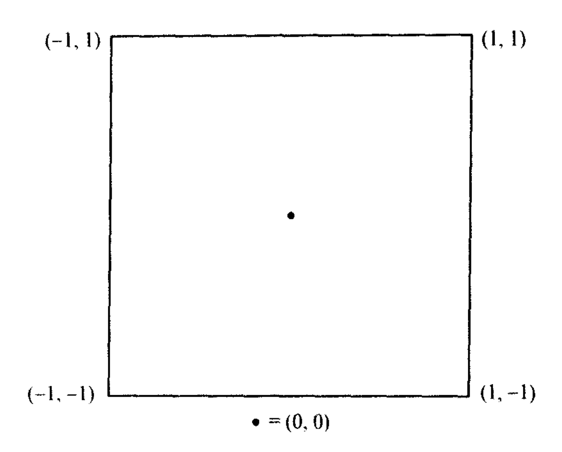


图 3-1　例 3a 的正方形区域

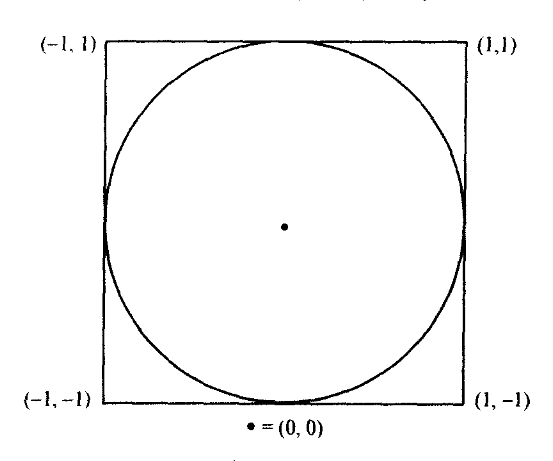


图 3-2　例 3a 的圆形区域

$$P\{(X,Y)\text{落在圆内}\} = P\{X^2+Y^2 \leqslant 1\} = \frac{\text{圆的面积}}{\text{正方形的面积}} = \frac{\pi}{4}$$

[44] 因此，如果我们生成大量的正方形内的随机点，则这些点落在此圆内的比例近似等于 $\pi/4$. 如果 X 和 Y 独立且均服从 $(-1,1)$ 上的均匀分布，则它们的联合密度为

$$f(x,y) = f(x)f(y) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

[45]

$$-1 \leqslant x \leqslant 1,\ -1 \leqslant y \leqslant 1$$

因为 (X,Y) 的密度函数在此正方形内是常数，故由定义知 (X,Y) 在此正方形上均

匀分布. 如果 U 在 $(0,1)$ 上均匀分布，则 $2U$ 在 $(0,2)$ 上均匀分布，$2U-1$ 在 $(-1,1)$ 上均匀分布. 因此，如果我们生成两个随机数 U_1 和 U_2，且令 $X=2U_1-1, Y=2U_2-1$,

$$I=\begin{cases}1 & 如果 X^2+Y^2\leqslant 1\\ 0 & 其他情况\end{cases}$$

则

$$E[I]=P\{X^2+Y^2\leqslant 1\}=\frac{\pi}{4}$$

这样，我们就可以如下估计 π：先生成大量的随机数对 (u_1,u_2)，之后用满足 $(2u_1-1)^2+(2u_2-1)^2\leqslant 1$ 的随机数对的比例来估计 $\pi/4$. □

于是，我们可以用随机数生成器来生成 $(0,1)$ 上均匀分布的随机变量. 我们将在第 4 章和第 5 章指出如何利用这些随机数来生成任一分布的随机变量值. 基于这一点，我们就可以模拟任一概率系统，也就是说，我们可以根据指定系统的概率分布去生成此系统的所有随机变量.

习 题

1. 如果 $x_0=5$，且

$$x_n=3x_{n-1} \bmod 150$$

求 $x_1,\cdots,x_{10}$.

2. 如果 $x_0=3$，

$$x_n=(5x_{n-1}+7) \bmod 200$$

求 $x_1,\cdots,x_{10}$.

应用随机模拟近似计算习题 3 至 习题 9 中的积分. 如果知道其精确答案，请比较二者的区别.

3. $\int_0^1 \exp\{e^x\}dx$

4. $\int_0^1 (1-x^2)^{3/2}dx$

5. $\int_{-2}^2 e^{x+x^2}dx$

6. $\int_0^\infty x(1+x^2)^{-2}dx$

7. $\int_{-\infty}^\infty e^{-x^2}dx$

8. $\int_0^1\int_0^1 e^{(x+y)^2}dydx$

9. $\int_0^\infty\int_0^x e^{-(x+y)}dydx$

[提示：设 $I_y(x)=\begin{cases}1 & 如果 y<x\\ 0 & 如果 y\geqslant x\end{cases}$ 并应用此函数使其积分等于 1，其中两个积分限均为 0 到 ∞.]

46

10. 应用模拟方法近似计算 $\mathrm{Cov}(U, \mathrm{e}^U)$, 其中 U 为 (0,1) 均匀分布的随机变量, 并与其精确答案比较.

11. 设 U 在 $(0,1)$ 上均匀分布, 应用模拟方法近似计算

    (a) $\mathrm{Corr}(U, \sqrt{1-U^2})$.   (b) $\mathrm{Corr}(U^2, \sqrt{1-U^2})$.

12. 对于 (0,1) 上均匀分布的随机变量 $U_1, U_2, \cdots$, 定义

$$N = \min\left\{n : \sum_{i=1}^{n} U_i > 1\right\}$$

即 N 为总和超过 1 的随机变量的最小个数.

(a) 通过生成 100 个 N 估计 $E[N]$.

(b) 通过生成 1000 个 N 估计 $E[N]$.

(c) 通过生成 10000 个 N 估计 $E[N]$.

(d) 你认为 $E[N]$ 等于多少?

13. 设 U_i, $i \geqslant 1$ 为随机数, 定义 N 为

$$N = \max\left\{n : \prod_{i=1}^{n} U_i \geqslant \mathrm{e}^{-3}\right\}$$

其中 $\prod_{i=1}^{0} U_i \equiv 1$.

(a) 用模拟方法求 $E[N]$.

47

(b) 用模拟方法求 $P\{N = i\}$, $i = 0, 1, 2, 3, 4, 5, 6$.

14. 令

$$x_n = 3x_{n-1} + 5x_{n-2} \bmod (100), \ n \geqslant 3$$

其中 $x_1 = 23, x_2 = 66$. 我们称序列 $u_n = x_n/100$, $n \geqslant 1$ 为 text 随机数序列. 求其前 14 个值.

参考文献

[1] Knuth, D., *The Art of Computer Programming*, Vol.2, 2nd ed., *Seminumerical Algorithms*. Addison-Wesley, Reading, MA, 2000.

[2] L'Ecuyer, P., *Random Numbers for Simulation*, *Commun. Assoc. Comput. Mach.* **33**, 1990.

[3] Marsaglia, G., *Random Numbers Fall Mainly in the Planes*, *Proc. Natl. Acad. Sci. USA*, **61**, 25–28, 1962.

[4] Marsaglia, G., *The Structure of Linear Congruential Sequences*, in *Applications of Number Theory to Numerical Analysis*,S. K. Zaremba, ed., Academic Press, London, 1972, pp. 249–255.

[5] Naylor, T., *Computer Simulation Techniques.* Wiley, New York, 1966.

[6] Ripley, B., *Stochastic Simulation.* Wiley, New York, 1986.

[7] von Neumann, J., "Various Techniques Used in Connection with Random Digits, 'Monte Carlo Method,' " *U. S. National Bureau of Standards Applied Mathematics Series,* No.12, 36–38, 1951.

48

第 4 章　离散随机变量的生成

4.1 逆变换法

假设我们希望生成一个概率分布函数为

$$P\{X=x_j\}=p_j,\ j=0,1,\cdots,\sum_j p_j=1$$

的离散随机变量 X. 为此，我们先生成一个随机数 U，即 U 在 (0,1) 上均匀分布，且令

$$X=\begin{cases} x_0 & \text{如果} U<p_0 \\ x_1 & \text{如果} p_0\leqslant U<p_0+p_1 \\ \vdots & \\ x_j & \text{如果} \sum_{i=0}^{j-1}p_i\leqslant U<\sum_{i=0}^{j}p_i \\ \vdots & \end{cases}$$

对于 $0<a<b<1$，由于 $P\{a\leqslant U<b\}=b-a$，故我们有

$$P\{X=x_j\}=P\left\{\sum_{i=0}^{j-1}p_i\leqslant U<\sum_{i=0}^{j}p_i\right\}=p_j$$

所以，X 的分布满足要求.

注 1　算法步骤可以写成

生成一个随机数 U
如果 $U<p_0$，则令 $X=x_0$ 且停止
如果 $U<p_0+p_1$，则令 $X=x_1$ 且停止
如果 $U<p_0+p_1+p_2$，则令 $X=x_2$ 且停止
$\vdots$

注 2　如果 $x_i,\ i\geqslant 0$ 是由小到大排列的，即 $x_0<x_1<x_2<\cdots$，且以 F 记 X 的分布函数，则 $F(x_k)=\sum_{i=0}^{k}p_i$，且

$$\text{如果 } F(x_{j-1})\leqslant U<F(x_j)\text{，则 } X \text{ 等于 } x_j$$

换言之，当生成一个随机数 U 之后，我们是通过 U 是否落在区间 $[F(x_{j-1}), F(x_j))$ 来确定 X 的值（或等价地通过求取 $F(U)$ 的逆）. 基于此，上述方法被称为离散的逆变换法. □

用上述方法来生成一个离散随机变量所需的时间与我们要搜索的区间个数成正比，于是，我们有必要以 p_j 的降序排列 X 的取值 x_j.

例 4a 如果我们想模拟分布列为

$$p_1 = 0.20, \quad p_2 = 0.15, \quad p_3 = 0.25, \quad p_4 = 0.40$$

的随机变量 X(其中 $p_j = P\{X = j\}$)，则我们首先生成一个随机数 U，之后如下进行：

如果 $U < 0.20$，则令 $X = 1$ 且停止
如果 $U < 0.35$，则令 $X = 2$ 且停止
如果 $U < 0.60$，则令 $X = 3$ 且停止
否则令 $X = 4$.

然而，下述方法更有效：

如果 $U < 0.40$，则令 $X = 4$ 且停止
如果 $U < 0.65$，则令 $X = 3$ 且停止
如果 $U < 0.85$，则令 $X = 1$ 且停止
否则令 $X = 2$. □

50

当所求的随机变量为离散均匀随机变量时，上述区间搜索是不必要的，也就是说，当我们想生成的随机变量 X 以相等的概率取值 $1, \cdots, n$ 中的任一个，即 $P\{X = j\} = 1/n$, $j = 1, \cdots, n$ 时，上述区间搜索是没有必要的. 事实上，由上述方法可知，如果以 U 记生成的随机数，则

$$\text{当 } \frac{j-1}{n} \leqslant U < \frac{j}{n} \text{ 时，} X = j$$

即，当 $j - 1 \leqslant nU < j$ 时，$X = j$，或等价地有

$$X = \text{Int}(nU) + 1$$

其中 $\text{Int}(x)$ 有时也写成 $[x]$，是 x 的整数部分，即不大于 x 的最大整数.

在随机模拟中，离散均匀随机变量是非常重要的，请参见下面的两个例子.

例 4b 随机排列的生成 假设我们感兴趣的是生成 $1, 2, \cdots, n$ 的一个随机排列，即它的 $n!$ 种排列是等可能的. 此随机排列可如下进行：先随机地在数 $1, 2, \cdots, n$ 中任选一个，并把此数放在位置 n 上；之后，在其余 $n - 1$ 个数中随机地任选一个，并把此数放在位置 $n - 1$ 上；再从其余 $n - 2$ 个数中随机地任选一个，并把此数放

在位置 $n-2$ 上；依此类推（随机选数意味着每个数被选中的概率是相同的）. 此种排列方法不必严格考虑哪一个数被排列在什么位置上. 然而，另一个更方便和有效的方法是这些数的随机排列是通过其位置的随机排列得到的. 即，如果这些数的初始顺序是 $P_1, P_2, \cdots, P_n$，则我们在位置 $1, 2, \cdots, n$ 中随机地任选一个，之后把此位置上的数与位置 n 上的数互换. 然后，从位置 $1, 2, \cdots, n-1$ 中随机地任选一个，再把此位置上的数与位置 $n-1$ 上的数互换，依此类推.

我们注意到，$\text{Int}(kU)+1$ 等可能地在 $1, 2, \cdots, k$ 中任取一值，于是，生成随机排列的上述算法可以通过如下步骤完成：

步骤 1：设 $P_1, P_2, \cdots, P_n$ 是 $1, 2, \cdots, n$ 的任一排列（例如可取 $P_j = j$, $j = 1, \cdots, n$）.

步骤 2：令 $k = n$.

步骤 3：生成一随机数 U，并设 $I = \text{Int}(kU) + 1$.

51

步骤 4：交换 P_I 和 P_k.

步骤 5：令 $k = k - 1$，如果 $k > 1$，则转至步骤 3.

步骤 6：$P_1, \cdots, P_n$ 为所求的随机排列.

例如，设 $n = 4$ 且初始排列为 1,2,3,4. 如果 I 的第一个值为 3(1,2,3 或 4 中的任一个均是等可能的)，则交换 3 和 4 上的数而得到新的排列为 1,2,4,3. 如果下一个 I 的值是 2，则交换位置 2 和 3 上的数而得到新的排列为 1,4,2,3. 如果 I 的最后一个值为 2，则最终的排列为 1,4,2,3，这即是所求的随机排列. □

上述算法的一个非常重要的性质是它可用来生成整数 $1, 2, \cdots, n$ 的容量为 r 的随机子集, 这就是说，应用上述算法直至位置 $n, n-1, \cdots, n-r+1$ 被填满为止，就生成了容量为 r 的随机子集 (此时我们假设 $r \leqslant n/2$; 如果 $r > n/2$，则我们可以通过选取容量为 $n-r$ 的随机子集之后，把其余的 r 个数看成所求的随机子集即可).

我们注意到，上述生成随机子集的方法在医学试验中非常重要. 例如，假设某医学中心计划检测一种用来降低病人血液中胆固醇含量的新药的疗效，现征集到 1000 名志愿者做此检验. 为了减少某些外在因素如天气变化对志愿者血液中胆固醇含量的影响，现决定把这些志愿者分成容量为 500 的两组，一组是服用药物的治疗组，另一组是服用安慰剂的控制组. 另外，志愿者及药厂的工作人员均不知道谁被分到哪一组（称这样的检测为双盲的）. 余下的问题是需要确定哪些志愿者被分到治疗组. 显然，我们希望除了第一组的人服用药物，第二组的人服用安慰剂这一点区别外，在所有其他方面，两组的志愿者尽可能地相似. 这样做的目的在于保证两组响应的区别完全是由药物引起的. 为更好地做到这一点，采取一个大家都认同的方法，从 1000 人中完全随机地选取 500 人构成治疗组，也就是说，从 1000 人选

500 人的 $\binom{1000}{500}$ 个子集中的任一个均是等可能的.

注 生成随机排列的另一种方法是先生成 n 个随机数 $U_1, \cdots, U_n$, 把它们排序后, 用其下标作为一个随机排列. 例如, 当 $n = 4, U_1 = 0.4, U_2 = 0.1, U_3 = 0.8, U_4 = 0.7$ 时, 由于 $U_2 < U_1 < U_4 < U_3$, 则随机排列为 2,1,4,3. 然而, 这种方法的一个难点在于 n 个随机数的排序需要的计算量为 $n\ln(n)$. □

52

例 4c 平均值的计算 假设我们希望近似计算 $\bar{a} = \sum_{i=1}^{n} a(i)/n$, 其中 n 非常大且 $a(i), i = 1, \cdots, n$ 的值非常复杂不易计算. 计算此平均值的一种方法是: 如以 X 记取值为 $1, 2, \cdots, n$ 的离散均匀随机变量, 则随机变量 $a(X)$ 的均值为

$$E[a(X)] = \sum_{i=1}^{n} a(i)P\{X = i\} = \sum_{i=1}^{n} \frac{a(i)}{n} = \bar{a}$$

如果我们生成 k 个离散均匀随机变量 $X_i, i = 1, \cdots, k$ (可以先生成 k 个随机数 U_i, 然后令 $X_i = \text{Int}(nU_i) + 1$), 则此 k 个随机变量 $a(X_i)$ 的均值为 $\bar{a}$. 于是, 由强大数定律可知, 当 k 足够大时 (它仍可比 n 小许多), 这些值的平均值将近似等于 $\bar{a}$. 因此, 我们有

$$\bar{a} \approx \sum_{i=1}^{k} \frac{a(X_i)}{k}$$

□

在生成几何随机变量时, 我们不需要搜索它所落入的区间.

例 4d 我们在前面讲过, 参数为 p 的几何随机变量 X 的分布律为

$$P\{X = i\} = pq^{i-1}, \ i \geqslant 1$$

其中 $q = 1 - p$, 并且 X 表示在成功概率为 p 的多次独立重复试验中成功首次出现时的试验次数. 因为

$$\begin{aligned}\sum_{i=1}^{j-1} P\{X = i\} &= 1 - P\{X > j - 1\} \\ &= 1 - P\{\text{前} j - 1 \text{ 次试验均失败}\} \\ &= 1 - q^{j-1}, \ j \geqslant 1\end{aligned}$$

于是, 我们可以通过生成随机数 U, 并令 X 等于满足下式的 j,

$$1 - q^{j-1} \leqslant U < 1 - q^j$$

或等价地

$$q^j < 1 - U \leqslant q^{j-1}$$

53

即，我们可以定义 X 为

$$X = \min\{j : q^j < 1 - U\}$$

因为对数函数是严格单调增的，故 $a < b$ 等价于 $\ln(a) < \ln(b)$，则 X 可表示为

$$\begin{aligned} X &= \min\{j : j\ln(q) < \ln(1-U)\} \\ &= \min\left\{j : j > \frac{\ln(1-U)}{\ln(q)}\right\} \end{aligned}$$

其中最后一个不等号变号的原因是对于在区间 (0,1) 内的 q, $\ln(q)$ 为负的. 于是，利用记号 Int()，我们可以把 X 表示成

$$X = \mathrm{Int}\left(\frac{\ln(1-U)}{\ln(q)}\right) + 1$$

最后，注意到 $1-U$ 是 (0,1) 上均匀分布的随机变量，故

$$X \equiv \mathrm{Int}\left(\frac{\ln(U)}{\ln(q)}\right) + 1$$

也是参数为 p 的几何随机变量. □

例 4e　独立伯努利随机变量序列的生成　假设想生成 n 个独立的参数为 p 的伯努利随机变量 $X_1, \cdots, X_n$. 此随机变量容易通过先生成 n 个随机数 $U_1, \cdots, U_n$，然后令

$$X_i = \begin{cases} 1 & \text{如果}U_i \leqslant p \\ 0 & \text{如果}U_i > p \end{cases}$$

而得到. 现在我们将给出一种效率更高的方法. 为此，我们设想这些随机变量表示一个随机试验列的结果，如 $X_i = 1$，则表示第 i 次试验成功，否则就表示失败. 为生成当 $p \leqslant 1/2$ 时的这些试验结果，我们应用例 4d 的结论去生成一个几何随机变量 N(它表示在每次成功概率为 p 的随机试验列中首次成功出现时的试验次数). 假设 N 的模拟值为 $N = j$. 如果 $j > n$，则令 $X_i = 0$, $i = 1, \cdots, n$; 如果 $j \leqslant n$，则令 $X_1 = \cdots = X_{j-1} = 0, X_j = 1$; 如果 $j < n$，则重复上面操作以得到其余 $n - j$ 个伯努利随机变量的值. (当 $p > 1/2$ 时，由于我们想同时生成尽可能多的伯努利随机变量，则我们应当生成首次失败而不是首次成功出现时的试验次数.)

54

当 X_i 是独立但不是同分布的伯努利随机变量时，我们仍可以应用上述思想. 对每个 $i = 1, \cdots, n$，令 u_i 为 X_i 最不可能取的两个值中的一个，即如果 $P\{X_i = 1\} \leqslant 1/2$，则 $u_i = 1$，否则 $u_i = 0$. 另外，记 $p_i = P\{X_i = u_i\}, q_i = 1 - p_i$. 我们将通过先生成随机变量 X 的值来模拟此伯努利序列：对于 $j = 1, \cdots, n$，如最不可能取

的值首次出现在第 j 次试验，则 X 等于 j; 如果在 n 次试验中没有出现最不可能取的值，则 X 等于 $n+1$. 为生成 X，设 $q_{n+1}=0$，且注意到

$$P\{X>j\}=\prod_{i=1}^{j} q_i,\ j=1,\cdots,n+1$$

于是，

$$P\{X \leqslant j\}=1-\prod_{i=1}^{j} q_i,\ j=1,\cdots,n+1$$

这样我们可以通过生成随机数 U，然后令

$$X=\min\left\{j: U \leqslant 1-\prod_{i=1}^{j} q_i\right\}$$

来模拟 X. 如果 $X=n+1$，则此伯努利随机变量的模拟序列为 $X_i=1-u_i, i=1,\cdots,n$. 如果 $X=j(j \leqslant n)$, 则令 $X_i=1-u_i, i=1,\cdots,j-1, X_j=u_j$; 如果 $j<n$，则用类似方法生成其余 $X_{j+1},\cdots,X_n$ 的值. □

4.2 泊松随机变量的生成

如果随机变量 X 满足

$$p_i=P\{X=i\}=\mathrm{e}^{-\lambda}\frac{\lambda^i}{i!},\quad i=0,1,\cdots$$

则它是参数为 λ 的泊松随机变量. 利用逆变换法生成上述泊松随机变量的关键是下面的等式（它在第 2 章的 2.4 节已经证明过）：

$$p_{i+1}=\frac{\lambda}{i+1}p_i,\quad i \geqslant 0 \tag{4.1}$$

基于上述计算泊松概率的递推公式，利用逆变换法生成参数为 λ 的泊松随机变量的算法可以如下表示，其中量 i 表示随机变量的取值; $p=p_i$ 表示 X 等于 i 的概率，$F=F(i)$ 是 X 小于等于 i 的概率.

55

步骤 1：生成一个随机数 U.

步骤 2：$i=0, p=\mathrm{e}^{\lambda}, F=p$.

步骤 3：如果 $U<F$，取 $X=i$ 且停止.

步骤 4：$p=\lambda p/(i+1), F=F+p, i=i+1$.

步骤 5：转向步骤 3.

在上面的算法中，我们注意到 $i=i+1$ 并不意味着 i 等于 $i+1$，而是表示将 i 的值增加 1. 为验证上述算法的确可以生成参数为 λ 的泊松随机变量，我们注意到它首先生成一个随机数 U，然后验证 $U<\mathrm{e}^{-\lambda}=p_0$ 是否成立. 如果上一不等式成立，则令 $X=0$；否则，利用 (4.1) 式计算第 4 步中的 p_1，并验证 $U<p_0+p_1$ 是否成立 (现在右端的 F 值是新的). 如果此不等式成立，则令 $X=1$. 依此类推.

上述算法逐步检验泊松变量的值是否为 0，然后是否为 1，是否为 2 等，且所需要的比较次数比生成的泊松随机变量的值大 1. 于是，上述算法的平均搜索次数为 $1+\lambda$. 故当 λ 比较小时，上述算法是合适的. 而当 λ 比较大时，上述算法还可以极大地改进. 事实上，因为参数为 λ 的泊松随机变量最可能取的值为两个最接近 λ 的整数之一，于是，一个更有效的算法应首先检验这两个整数，而不是从初值 0 开始且向上检测. 例如，记 $I=\mathrm{Int}(\lambda)$ 且用等式 (4.1) 逐步确定 $F(I)$. 现通过生成一个随机数 U 来产生一个参数为 λ 的泊松随机变量 X. X 是否小于等于 I 由 U 是否小于等于 $F(I)$ 来决定. 如果 $X\leqslant I$，则从 I 开始向下搜索；否则从 $I+1$ 开始向上搜索.

这种算法所需的搜索次数比随机变量 X 与其均值 λ 的绝对差大约多 1. 因为对于较大的 λ，由中心极限定理可知，泊松随机变量近似于均值和方差均为 λ 的正态随机变量，故

$$
\begin{aligned}
\text{平均搜索次数} &\simeq 1+E[|X-\lambda|] && \text{其中}X\sim N(\lambda,\lambda)\\
&=1+\sqrt{\lambda}E\left[\frac{|X-\lambda|}{\sqrt{\lambda}}\right] &&\\
&=1+\sqrt{\lambda}E[|Z|] && \text{其中}Z\sim N(0,1)\\
&=1+0.798\sqrt{\lambda} && \text{见习题 11}
\end{aligned}
$$

其中记号 $X\sim F$ 表示 X 的分布函数为 F，$N(\mu,\sigma^2)$ 表示均值为 μ，方差为 σ^2 的正态分布. 从上式可以看出，当 λ 增大时，算法 4-1 的平均搜索次数的增长速度等同于 λ 的平方根，而不是 λ 本身.

56

4.3　二项随机变量的生成

假设我们希望生成参数为 (n,p) 的二项随机变量 X，此时 X 的分布为

$$P\{X=i\}=\frac{n!}{i!(n-i)!}p^i(1-p)^{n-i},\ i=0,1,\cdots,n$$

基于下面的递推等式

$$P\{X=i+1\}=\frac{n-i}{i+1}\frac{p}{1-p}P\{X=i\}$$

我们研究如何利用逆变换法来产生上述二项分布的随机变量.

用 i 记随机变量的当前取值, $pr = P\{X = i\}$ 记 X 等于 i 的概率, $F = F(i)$ 记 X 小于等于 i 的概率. 则算法可以如下描述:

生成二项随机变量 $B(n, p)$ 的逆变换算法

步骤 1: 生成一个随机数 U.

步骤 2: $c = p/(1-p), i = 0, pr = (1-p)^n, F = pr$.

步骤 3: 如果 $U < F$, 令 $X = i$ 并停止.

步骤 4: $pr = [c(n-i)/(i+1)]pr, F = F + pr, i = i + 1$.

步骤 5: 转至步骤 3.

上述算法首先检验 $X = 0$ 成立与否, 之后检验 $X = 1$ 成立与否等. 因此, 此算法的搜索次数比 X 的取值多 1, 即生成 X 的平均搜索次数为 $1 + np$. 由于二项随机变量 $B(n,p)$ 为在 n 重独立的成功概率为 p 的试验中成功出现的次数, 故此随机变量也可用 n 减去一个二项随机变量 $B(n, 1-p)$ 来得到 (为什么如此?). 于是, 当 $p > \frac{1}{2}$ 时, 我们可以先生成一个二项随机变量 $B(n, 1-p)$, 然后用 n 减去此随机变量而得到所求的二项随机变量.

注 1 另一个产生二项随机变量 $B(n,p)$ 的方法是: 利用二项随机变量为 n 重独立的每次成功概率为 p 的伯努利试验中成功的次数的解释. 同样, 我们也可以通过生成 n 个伯努利试验的结果来模拟 X.

57

注 2 如同产生泊松随机变量, 当其均值 np 很大时, 最好是先判断生成的随机变量值是否小于等于 $I \equiv \text{Int}(np)$ 或大于 I. 对于第一种情况, 我们应从 I 开始搜索, 然后是 $I-1$, $I-2, \cdots$ 等等; 对于第二种情况, 我们应从 $I+1$ 开始向上搜索.

4.4 筛选技术

假设我们有一种生成概率分布函数为 $\{q_j, j \geqslant 0\}$ 的随机变量的有效方法. 基于这种方法, 我们可以模拟一个分布函数为 $\{p_j, j \geqslant 0\}$ 的随机变量, 其方法是先模拟一个分布函数为 $\{q_j\}$ 的随机变量 Y, 然后以正比例于 p_Y/q_Y 的概率接受这个模拟值.

特别地, 设 c 为一常数且对于所有满足 $p_j > 0$ 的 j 有

$$\frac{p_j}{q_j} \leqslant c \tag{4.2}$$

于是, 我们给出下面的舍弃法或称筛选法以模拟分布函数为 $p_j = P\{X = j\}$ 的随机变量 X.

筛选法

步骤 1：模拟一个概率分布函数为 q_j 的随机变量值 Y.

步骤 2：生成一个随机数 U.

步骤 3：如果 $U < p_Y/cq_Y$，则令 $X = Y$ 且停止. 否则，转至步骤 1.

上述方法的图示解释请见图 4-1.

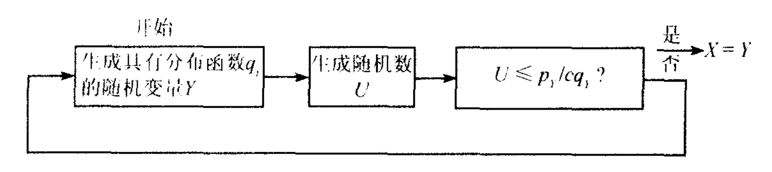


图 4-1　筛选法

下面我们证明上述筛选方法是正确的.

定理　筛选算法生成的随机变量 X 的分布函数为

58

$$P\{X = j\} = p_j,\ j = 0, 1, \cdots$$

另外，得到随机变量 X 的上述算法的迭代次数是一个参数为 c 的几何随机变量.

证明　让我们先求一次迭代所生成接受值 j 的概率. 注意到

$$\begin{aligned} P\{Y = j, 且被接受\} &= P\{Y = j\}P\{接受此值|Y = j\} \\ &= q_j \frac{p_j}{cq_j} = \frac{p_j}{c} \end{aligned}$$

对上式两边关于 j 求和后，则得到生成的随机变量被接受的概率为

$$P\{此值被接收\} = \sum_j \frac{p_j}{c} = \frac{1}{c}$$

因为每次迭代都是独立的且以概率 $1/c$ 产生一个接收者，故所需的迭代次数是均值为 c 的几何随机变量，另外，

$$\begin{aligned} P\{X = j\} &= \sum_n P\{j在第n次迭代被接受\} \\ &= \sum_n (1 - 1/c)^{n-1} \frac{p_j}{c} = p_j \end{aligned}$$

□

注　读者请注意，以概率 p_Y/cq_Y 接受 Y 的值的含义是：生成一个随机数 U 后，如果 $U \leqslant p_Y/cq_Y$，则接收 Y.　□

例 4f 假设我们希望模拟一个以概率 0.11, 0.12, 0.09, 0.08, 0.12, 0.10, 0.09, 0.09, 0.10,0.10 分别取值 $1, 2, \cdots, 10$ 的随机变量 X. 一种方法是利用逆变换法，另一种方法是应用以概率 q 取值 $1, 2, \cdots, 10$ 的筛选法. 此时，$q_j = 1/10, j = 1, 2, \cdots, 10$，且我们取 c 为

$$c = \max \frac{p_j}{q_j} = 1.2$$

其算法如下：

步骤 1：生成一个随机数 U_1，并令 $Y = \text{Int}(10U_1) + 1$.

步骤 2：生成第二个随机数 U_2.

步骤 3：如果 $U_2 \leqslant p_Y/0.12$，则令 $X = Y$ 且停止；否则转至步骤 1.

59

由于 $cq_Y = 1.2/10 = 0.12$，则步骤 3 中的常数为 0.12. 为生成 X 的值，上述算法所需的平均迭代次数为 1.2. □

此筛选法最早由著名数学家约翰・冯诺依曼提出，当我们用类似的筛选法来产生连续随机变量时，其功效会更加显著.

4.5 复合法

假设有一种有效方法可以用来生成概率分布函数为 $\{p_j^{(1)}, j \geqslant 0\}$ 或 $\{p_j^{(2)}, j \geqslant 0\}$ 的随机变量，并且我们想模拟概率分布函数为

$$P\{X = j\} = \alpha p_j^{(1)} + (1-\alpha) p_j^{(2)}, \ j \geqslant 0 \tag{4.3}$$

的随机变量 X，其中 $0 < \alpha < 1$. 模拟此随机变量 X 的一个方法是：如果 X_1, X_2 是概率分布分别为 $\{p_j^{(1)}\}$ 和 $\{p_j^{(2)}\}$ 的随机变量，则随机变量

$$X = \begin{cases} X_1 & \text{取值概率为}\alpha \\ X_2 & \text{取值概率为}1-\alpha \end{cases}$$

的分布函数由等式 (4.3) 给出. 由此可知，我们可以如下生成随机变量 X 的值：生成随机数 U 后，如果 $U < \alpha$，则生成 X_1 的值；如果 $U > \alpha$，则生成 X_2 的值.

例 4g 假设我们希望生成概率分布为

$$p_j = P\{X = j\} = \begin{cases} 0.05 & j = 1, 2, 3, 4, 5 \\ 0.15 & j = 6, 7, 8, 9, 10 \end{cases}$$

的随机变量 X. 由于 $p_j = 0.5p_j^{(1)} + 0.5p_j^{(2)}$，其中

$$p_j^{(1)} = 0.1, \ j = 1, \cdots, 10 \qquad p_j^{(2)} = \begin{cases} 0 & j = 1, 2, 3, 4, 5 \\ 0.2 & j = 6, 7, 8, 9, 10 \end{cases}$$

60

于是, X 的模拟值可以如下求得：生成一个随机数 U 后，如果 $U < 0.5$，则所求值即为从取值 $1,\cdots,10$ 的离散均匀分布抽取的随机变量值；否则，所求值即为从取值 $6,\cdots,10$ 的离散均匀分布抽取的随机变量值. 这样，我们可以模拟 X 如下：

步骤 1：生成一个随机数 U_1.

步骤 2：生成另一个随机数 U_2.

步骤 3：如果 $U_1 < 0.5$，则令 $X=\text{Int}(10U_2)+1$；否则，令 $X=\text{Int}(5U_2)+6$. □

如果 F_i, $i=1,\cdots,n$ 为分布函数，α_i, $i=1,\cdots,n$ 为非负常数且其和为 1，则称分布函数

$$F(x)=\sum_{i=1}^{n}\alpha_i F_i(x)$$

为 F_i, $i=1,\cdots,n$ 的混合或复合分布函数. 模拟来自混合分布函数 F 的随机变量的一个方法是：先模拟随机变量 I(其取 i 的概率为 $\alpha_i, i=1,\cdots,n$)，然后模拟来自分布函数 F_I 的随机变量（如果 I 的模拟值为 j，则模拟来自 F_j 的随机变量）. 我们称这种方法为复合法.

4.6 随机向量的生成

随机向量 $X_1,\cdots,X_n$ 的模拟可以通过逐步生成 X_i 来完成，即先生成 X_1，再从给定 X_1 的生成值下 X_2 的条件分布生成 X_2，从给定 X_1,X_2 的生成值下 X_3 的条件分布生成 X_3，等等. 例 4g 以如何模拟多项分布为例对此方法进行说明.

例 4h　考虑 n 次独立试验，其每次试验的结果均为 $1,2,\cdots,r$ 中的一个，且其相应概率分别为 $p_1,p_2,\cdots,p_r(\sum_{i=1}^{r}p_i=1)$. 如以 X_i 记结果为 i 的试验次数，则称随机向量 $(X_1,\cdots,X_r)$ 为多项随机向量, 其联合概率分布函数为

$$P\{X_i=x_i,\ i=1,\cdots,r\}=\frac{n!}{x_1!\cdots x_r!}p_1^{x_1}\cdots p_r^{x_r}$$

其中 $\sum_{i=1}^{r}x_i=n$.

模拟上述随机向量的最好方法依赖于 r 与 n 的大小. 如果相对于 n，r 很大，则在任一次试验中，许多试验结果都不会出现. 于是，通过生成 n 次试验结果来模拟这样的随机向量也许是最好的方法，即先独立地生成随机变量 $Y_1,\cdots,Y_n$，其中

61

$$P\{Y_j=i\}=p_i,\ i=1,\cdots,r,\ j=1,\cdots,n$$

然后，令

$$X_i=\#\{j:Y_j=i,j=1,\cdots,n\}$$

其中 # 表示此集合中的元素个数. 从上述方法可以看出, Y_j 的生成值表示第 j 次试验的结果, 而 X_i 则是试验结果为 i 的试验次数.

另一方面, 如果相对于 r, n 很大, 则可以依次模拟 $X_1,\cdots,X_n$ 以产生需要的随机向量, 即先产生 X_1, 再产生 X_2, 等等. 由于 n 次试验是独立的, 且均以概率 p_1 出现结果 1, 故 X_1 是一个参数为 (n,p_1) 的二项随机变量. 于是, 我们可以利用 4.3 节的方法模拟 X_1, 并假设 x_1 为其生成值. 当已知在 n 次试验中结果 1 出现 x_1 次时, 由于其他 $n-x_1$ 次试验是独立的, 则结果 2 出现的概率为

$$P\{\text{结果 2 出现}\mid\text{结果 1不会出现}\}=\frac{p_2}{1-p_1}$$

故由此可知, 给定 $X_1=x_1$ 下, X_2 的条件分布是参数为 $(n-x_1,\frac{p_2}{1-p_1})$ 的二项分布. 同样, 我们可以利用 4.3 节中的方法生成 X_2 的值, 并记其值为 x_2. 然后在 $X_1=x_1,X_2=x_2$ 的条件下生成 X_3 的值. 此时由于在 n 次试验中有 x_1 次试验结果为 1, 有 x_2 次试验结果为 2, 故在剩余的 $n-x_1-x_2$ 次独立试验中, 试验结果 3 出现的概率为 $\frac{p_3}{1-p_1-p_2}$. 于是, 给定 $X_i=x_i,i=1,2$ 下 X_3 的条件分布为参数 $(n-x_1-x_2,\frac{p_3}{1-p_1-p_2})$ 的二项分布, 并由此可以产生 X_3. 依此类推, 则得到了 $X_1,\cdots,X_r$ 的模拟值. □

习 题

1. 编写生成 n 个来自概率分布为 $p_1=1/3,p_2=2/3$ 的模拟值的程序.
   (a) 设 $n=100$, 运行此程序, 且求它们等于 1 的比例.
   (b) 重复当 $n=1000$ 时的 (a).
   (c) 重复当 $n=10\ 000$ 时的 (a).
2. 编写如下计算机程序: 输入是概率分布函数为 $\{p_j,\ j=1,\cdots,n\}$ 的随机变量, 输出是与此分布相同的随机变量的值.
3. 给出一个有效算法以模拟分布为

$$P\{X=1\}=0.3,\ P\{X=2\}=0.2,\ P\{X=3\}=0.35,\ P\{X=4\}=0.15$$

   的随机变量的值.

62

4. 洗好一副编号分别为 $1,2,\cdots,100$ 的纸牌, 现每次从中任选一张, 如果第 i 次抽中的纸牌正好为纸牌 i, 则称之为成功 $(i=1,\cdots,100)$. 编写程序估计成功次数的期望与方差, 并运行此程序. 之后, 计算其严格答案, 并与你的模拟估计进行比较.
5. 一种有别于例 4b 的生成随机排列的方法是: 依次从 $n=1$, $n=2$, $n=3,\cdots$ 模拟产生元素 $1,2,\cdots,n$ 的一个随机排列.(显然, 当 $n=1$ 时的随机排列就是 1. 如以

$P_1, \cdots, P_{n-1}$ 记前 $n-1$ 个元素的随机排列，则 n 个元素 $1, 2, \cdots, n$ 的随机排列如下得到：先从 $P_1, \cdots, P_{n-1}, n$ 中等可能地任取一个位置，之后交换此位置与位置 n 上的数.)

(a) 编写上述方法的程序.

(b) 用数学归纳法证明上述方法可行，即 $1, 2, \cdots, n$ 的 $n!$ 个排列均等可能地由本方法得到.

6. 与 text 随机数序列一起应用一种有效方法生成 25 个独立的参数为 $p = 0.8$ 的伯努利随机变量. 需要多少个随机数？

7. 连续投掷两枚均匀骰子直到其点数之和的所有结果 $2, 3, \cdots, 12$ 均至少出现一次. 请给出估计投掷次数的期望与方差的模拟方法.

8. 假设某清单共有 n 项，且其中任一项均有一数值与之对应，比如，以数 $v(i)$ 对应着第 i 项. 再假设 n 非常大，且每一项均可以出现在清单的许多不同位置上. 请解释如何利用随机数来估计此清单不同项目所对应的数字之和. (尽管某项可能多次出现在此清单上，但求和时仅计算一次.)

9. 对于 n 个事件 $A_1, \cdots, A_n$，其中 A_i 包含 n_i 个试验结果：$A_i = \{a_{i,1}, a_{i,2}, \cdots, a_{i,n_i}\}$. 对于任一给定的试验结果 a，假设其出现的概率 $P\{a\}$ 已知. 请问如何利用习题 8 的结果估计 $P\{\cup_{i=1}^{n} A_i\}$，即事件 A_i 中至少有一个出现的概率 (注意我们并没有假设事件 A_i, $i = 1, \cdots, n$ 互斥).

10. 参数为 (r, p) 的负二项分布的概率分布函数为

63

$$p_j = \frac{(j-1)!}{(j-r)!(r-1)!} p^r (1-p)^{j-r}, \ j = r, r+1, \cdots$$

其中 r 为正整数，$0 < p < 1$.

(a) 利用负二项和几何随机变量间的关系和例 4d 的结果给出模拟负二项随机变量的算法.

(b) 验证

$$p_{j+1} = \frac{j(1-p)}{j+1-r} p_j$$

(c) 利用 (b) 中的关系式给出模拟负二项随机变量的第二种算法.

(d) 负二项随机变量是在成功概率 p 的独立重复试验中直到 r 次成功出现时的总试验次数，利用这一解释，给出模拟它的另一方法.

11. 如果 Z 是一标准正态随机变量，证明

$$E[|Z|] = \left(\frac{2}{\pi}\right)^{1/2} \approx 0.798$$

12. 设随机变量 X 的分布为

$$P\{X=i\}=\frac{\mathrm{e}^{-\lambda}\lambda^i/i!}{\sum\limits_{j=0}^{k}\mathrm{e}^{-\lambda}\lambda^j/j!},\ i=0,1,\cdots,k$$

给出产生此随机变量 X 的两种方法.

13. 设 X 是参数为 (n,p) 的二项随机变量，且对于某个 $k\leqslant n$，随机变量 Y 的分布函数为给定 $X\geqslant k$ 下 X 的条件分布函数. 我们希望模拟随机变量 Y 的值. 记 $\alpha=P\{X\geqslant k\}$ 且假设 α 已求出.

(a) 给出模拟 Y 的逆变换法.

(b) 给出生成 Y 的第二种算法.

(c) α 取值大还是小，第二种算法无效？

14. 设随机变量 X 的概率分布函数为 p_j，$j=5,6,\cdots,14$，其中

$$p_j=\begin{cases}0.11 & \text{当}j\text{为奇数且 } 5\leqslant j\leqslant 13\\ 0.09 & \text{当}j\text{为偶数且 } 6\leqslant j\leqslant 14\end{cases}$$

给出模拟此随机变量的方法.

64

15. 设随机变量 X 分别以概率 0.06, 0.06, 0.06, 0.06, 0.06, 0.15, 0.13, 0.14, 0.15, 0.13 取值 $1,2,\cdots,10$. 应用复合方法给出产生此随机变量的算法.

16. 设随机变量 X 的分布为

$$P\{X=j\}=\left(\frac{1}{2}\right)^{j+1}+\frac{\left(\frac{1}{2}\right)2^{j-1}}{3^j},\ j=1,2,\cdots$$

给出模拟此随机变量的方法.

17. 设随机变量 X 的分布函数为 $p_j=P\{X=j\}$，其中 $\sum_{j=1}^{\infty}p_j=1$. 记

$$\lambda_n=P\{X=n|X>n-1\}=\frac{p_n}{1-\sum\limits_{j=1}^{n-1}p_j},\ n=1,2,\cdots$$

(a) 证明 $p_1=\lambda_1, p_n=(1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_{n-1})\lambda_n$. 如果我们把 X 看作某产品的寿命，则 λ_n 表示此产品在其寿命大于 $n-1$ 的条件下，寿命为 n 的概率. 于是，λ_n，$n\geqslant 1$ 被称为离散机会比. 下面一个被称为离散机会比方法的模拟离散随机变量的方法是：连续生成随机数直至第 n 个小于 λ_n. 其算法如下：

步骤 1：$X=1$.

步骤 2：生成随机数 U.

步骤 3：如果 $U<\lambda_X$，则停止.

步骤 4：$X=X+1$.

步骤 5：转至步骤 2.

(b) 证明上述算法产生的 X 具有已知的分布函数.

(c) 假设 X 是一个参数为 p 的几何随机变量，求 λ_n，$n \geqslant 1$. 说明此时的上述算法是用来做什么的，它的有效性为什么是明显的.

18. 假设对所有的 $n \geqslant 1$，$0 \leqslant \lambda_n \leqslant \lambda$. 考虑利用离散机会比 λ_n 产生随机变量的下述算法：

步骤 1：$S = 0$.

65

步骤 2：产生 U 且令 $Y = \text{Int}\left(\frac{\ln(U)}{\ln(1-\lambda)}\right) + 1$.

步骤 3：$S = S + Y$.

步骤 4：产生 U.

步骤 5：如果 $U \leqslant \lambda_S/\lambda$，则令 $X = S$ 且停止. 否则，转至步骤 2.

(a) 步骤 2 中 Y 的分布是什么?

(b) 说明这一算法是做什么的.

(c) 证明 X 是一个离散机会比为 $\{\lambda_n\}$ 的随机变量.

19. 在一个罐中有 n 个球，其中 n_i 个为颜色 i，$i = 1, \cdots, r(\sum_{i=1}^{r} n_i = n)$，现从中随机地选 m 个球. 以 X_i 记所抽取球中颜色为 i 的球的个数，讨论模拟 $X_1, \cdots, X_r$ 的有效算法.

66

第 5 章　连续随机变量的产生

引　言

对于连续随机变量的模拟，都有类似于产生离散随机变量的各种方法与之对应. 生成连续随机变量的逆变换法和筛选法将在 5.1 节和 5.2 节介绍. 在 5.3 节，我们考虑生成正态随机变量的有效方法，即，极坐标法. 最后，在 5.4 节和 5.5 节，我们考虑泊松过程和非齐次泊松过程的模拟问题.

5.1　逆变换法

设 F 为某连续随机变量的分布函数. 我们将基于如下命题给出生成此随机变量的逆变换法.

命题　*设 U 是 (0,1) 上均匀分布的随机变量, 对于任一连续分布函数 F, 随机变量*

$$X = F^{-1}(U)$$

的分布函数为 F. [$F^{-1}(u)$ 被定义为满足 $F(x) = u$ 的 x.]

证明　以 F_X 记 $X = F^{-1}(U)$ 的分布函数，则

$$F_X(x) = P\{X \leqslant x\} = P\{F^{-1}(U) \leqslant x\} \tag{5.1}$$

由于 $F(x)$ 是一分布函数，故它是 x 的单调增函数，且不等式 “$a \leqslant b$” 等价于不等式 “$F(a) \leqslant F(b)$”. 于是，由等式 (5.1) 知

$$\begin{aligned} F_X(x) &= P\{F(F^{-1}(U)) \leqslant F(x)\} \\ &= P\{U \leqslant F(x)\} \qquad \text{因为} F(F^{-1}(U)) = U \\ &= F(x) \qquad \text{因为} U \text{在 (0,1) 上均匀分布} \end{aligned}$$

□

由上一命题知，如果随机变量 X 的分布函数 $F(x)$ 连续，则我们可以如下模拟它：生成随机数 U 后，取 $X = F^{-1}(U)$.

例 5a　如果随机变量 X 的分布函数为

$$F(x) = x^n,\ 0 < x < 1$$

则用上述方法模拟生成 X.

如果记 $x = F^{-1}(u)$, 则

$$u = F(x) = x^n$$

或等价地有

$$x = u^{1/n}$$

于是, 我们可以通过生成随机数 U 后, 令 $X = U^{1/n}$ 来得到随机变量 X 的模拟值. □

逆变换方法是一种非常好的产生指数随机变量的方法, 参见下面的例子.

例 5b　设 X 是参数为 1 的指数随机变量, 则其分布函数为

$$F(x) = 1 - \mathrm{e}^{-x}$$

如果设 $x = F^{-1}(u)$, 则

$$u = F(x) = 1 - \mathrm{e}^{-x}$$

或

$$1 - u = \mathrm{e}^{-x}$$

或取对数后有

$$x = -\ln(1 - u)$$

于是, 参数为 1 的指数分布的随机变量 X 可由下式产生:

$$X = F^{-1}(U) = -\ln(1 - U)$$

其中 U 为随机数.

由于 $1 - U$ 也是 (0,1) 上均匀分布的随机变量, 故 $-\ln(1 - U)$ 与 $-\ln(U)$ 同分布. 于是, 由上式可知, 一个随机数的负对数是参数为 1 的指数随机变量. 利用这一点进行随机模拟可以节省一些计算时间.

另外, 我们注意到, 如果 X 是均值为 1 的指数随机变量, 则对于任一正常数 c, cX 是均值为 c 的指数随机变量. 于是, 参数为 λ(均值为 $1/\lambda$) 的指数随机变量 X 可以由

$$X = -\frac{1}{\lambda}\ln U$$

产生, 其中 U 为随机数. □

注　上述方法也可用来生成泊松随机变量. 前面讲过, 当两个相继事件的间隔时间是独立的参数为 λ 的指数随机变量时, 这些事件就构成一个参数为 λ 的泊松过程 (见第 2 章的 2.9 节). 由此可知, 到时间 1 出现的事件数 $N(1)$ 服从参数为 λ

的泊松分布. 如果以 X_i, $i=1,2,\cdots$ 记相继两个事件 (第 i 个与第 $i-1$ 个事件) 的间隔时间, 则第 n 个事件发生的时间为 $\sum_{i=1}^{n} X_i$. 于是, 到时刻 1 发生的事件数可以表示为

$$N(1)=\max\left\{n:\sum_{i=1}^{n} X_i \leqslant 1\right\}$$

这就是说, 到时刻 1 发生的事件数等于时刻 1 时发生的事件序数的最大值 (例如, 如果时刻 1 时第 4 个事件发生, 而第 5 个事件没有发生, 故时刻 1 时总共有 4 个事件发生). 于是, 利用例 5b 的结果, 我们可以证明参数为 λ 的泊松随机变量 $N=N(1)$ 满足

$$\begin{aligned}N&=\max\left\{n:\sum_{i=1}^{n}-\frac{1}{\lambda}\ln U_i\leqslant 1\right\}\\&=\max\left\{n:\sum_{i=1}^{n}\ln U_i\geqslant -\lambda\right\}\\&=\max\{n:\ln(U_1\cdots U_n)\geqslant -\lambda\}\\&=\max\{n:U_1\cdots U_n\geqslant \mathrm{e}^{-\lambda}\}\end{aligned}$$

其中 $U_1,\cdots,U_n,\cdots$ 为随机数. 于是, 参数为 λ 的泊松随机变量 N 可以如下产生: 相继生成随机数直至它们的乘积小于 $\mathrm{e}^{-\lambda}$, 则 N 等于这些随机数的个数减 1, 即

$$N=\min\{n:U_1\cdots U_n<\mathrm{e}^{-\lambda}\}-1$$

□

基于伽玛分布和指数分布间的关系, 我们可以有效利用例 5b 的结果来生成 $\Gamma(n,\lambda)$ 随机变量.

例 5c 假设我们希望产生 $\Gamma(n,\lambda)$ 的随机变量. 因为此随机变量的分布函数 F 为

$$F(x)=\int_0^x \frac{\lambda \mathrm{e}^{-\lambda y}(\lambda y)^{n-1}}{(n-1)!}\mathrm{d}y$$

故其逆函数不可能显示地表出. 但是, 由于 $\Gamma\ (n,\lambda)$ 随机变量 X 可以表示成 n 个独立的参数为 λ 的指数随机变量的和 (见第 2 章的 2.3 节), 故我们可以利用例 5b 的结果来生成 X. 特别地, 参数为 (n,λ) 的伽玛随机变量 X 可以如下产生:

$$\begin{aligned}X&=-\frac{1}{\lambda}\ln U_1-\cdots-\frac{1}{\lambda}\ln U_n\\&=-\frac{1}{\lambda}\ln(U_1\cdots U_n)\end{aligned}$$

其中 $U_1,\cdots,U_n$ 为 n 个随机数. 等式 $\sum_{i=1}^{n}\ln x_i=\ln(x_1\cdots x_n)$ 的应用可以节省一些计算时间, 其原因是右端仅需一次对数计算, 而左端却需要 n 次对数计算. □

我们可以利用例 5c 的方法来有效地生成一组指数随机变量: 先生成它们和的随机变量值, 之后再生成每个随机变量在其和给定条件下的单个值. 例如, 为了生成两个独立同分布的参数为 1 的指数随机变量 X 和 Y, 我们先生成其和 $X+Y$ 的值, 并记 $t=X+Y$, 然后, 再利用第 2 章习题 36 的结论: 在给定 $X+Y=t$ 下, X 的条件分布是 $(0,t)$ 上的均匀分布. 基于上述思想, 生成一对参数为 1 的指数随机变量的算法如下:

步骤 1: 生成随机数 U_1 和 U_2.

步骤 2: 令 $t=-\ln(U_1U_2)$.

步骤 3: 生成随机数 U_3.

70

步骤 4: $X=tU_3, Y=t-X$.

若利用例 5c 的方法直接生成, 则 $X=-\ln U_1, Y=-\ln U_2$. 比较二者的异同, 我们知道上面的算法比直接算法少了一次对数运算, 但多了两次乘法和多生成了一个随机数.

我们也可以运用上述算法生成 k 个独立的参数为 1 的指数随机变量: 先利用 $-\ln(U_1\cdots U_k)$ 生成它们的和, 且记 $t=-\ln(U_1\cdots U_k)$; 之后, 再生成 $k-1$ 个新的随机数, 且把它们由小到大排列成 $U_{(1)}<U_{(2)}<\cdots<U_{(k-1)}$; 最后, 所求的 k 个指数随机变量为

$$t[U_{(i)}-U_{(i-1)}],\ i=1,2,\cdots,k$$

其中 $U_{(0)}\equiv 0, U_{(k)}\equiv 1$.

5.2 筛 选 法

假设我们已有一种生成密度函数为 $g(x)$ 的随机变量的方法, 现在将利用这种方法生成密度函数为 $f(x)$ 的随机变量. 为此, 我们先生成来自 g 的随机变量 Y, 然后以正比例于 $f(Y)/g(Y)$ 的概率接受此值.

特别地, 设 c 为一常数且满足

$$\frac{f(y)}{g(y)}\leqslant c,\ 对任意\ y$$

于是, 我们有如下的生成密度函数为 f 的随机变量的方法 (其图示见图 5-1).

筛选法

步骤 1: 生成密度函数为 g 的随机变量 Y.

步骤 2：生成随机数 U.

步骤 3：如果 $U \leqslant \frac{f(Y)}{cg(Y)}$，则令 $X=Y$. 否则，转至步骤 1.

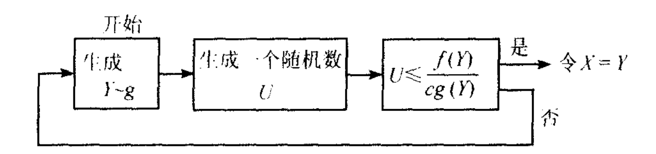


图 5-1 生成密度函数为 f 的随机变量的筛选法

71

有些读者会注意到，除了用密度函数替代概率分布函数外，产生连续随机变量的筛选法与离散情形完全一样. 于是，我们可以利用完全类似于离散情形的方法证明下面的结论.

定理

(i) 利用筛选法生成的随机变量的密度函数为 f.

(ii) 此算法所需的迭代次数是均值为 c 的几何随机变量.

与离散情形一样，以概率 $f(Y)/cg(Y)$ 接受 Y 可以如下进行：产生一个随机数 U 后，如果 $U \leqslant f(Y)/cg(Y)$，则接受 Y.

例 5d 用筛选法生成密度函数为

$$f(x)=20x(1-x)^3,\ 0<x<1$$

的随机变量.

由于此随机变量 [其分布为参数 (2,4) 的贝塔分布] 集中在区间 (0,1) 里，于是，我们选用

$$g(x)=1,\ 0<x<1$$

的筛选法. 为求取满足 $f(x)/g(x) \leqslant c$ 的最小的 c，我们利用微分法以确定

$$\frac{f(x)}{g(x)}=20x(1-x)^3$$

的最大值. 上式两边关于 x 求导，有

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{g(x)}\right)=20[(1-x)^3-3x(1-x)^2]$$

令上式等于零，则得到其最大值在点 $x=1/4$ 处达到. 于是

$$\frac{f(x)}{g(x)} \leqslant 20\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^3=\frac{135}{64} \equiv c$$

于是，

72

$$\frac{f(x)}{cg(x)} = \frac{256}{27}x(1-x)^3$$

这样，其筛选法算法如下：

步骤 1：生成随机数 U_1 和 U_2.

步骤 2：如果 $U_2 \leqslant \frac{256}{27}U_1(1-U_1)^3$，则令 $X = U_1$ 且停止. 否则，转至步骤 1.

步骤 1 的平均运行次数是 $c = \frac{135}{64} \approx 2.11$. □

例 5e 假设我们希望产生密度函数为

$$f(x) = Kx^{1/2}\mathrm{e}^{-x},\ x > 0$$

的 $\Gamma(3/2,1)$ 的随机变量，其中 $K = 1/\Gamma(3/2) = 2/\sqrt{\pi}$. 由于此随机变量的取值为正且均值为 3/2，故我们自然想到采用均值也为 3/2 的指数随机变量的密度作为 $g(x)$，即

$$g(x) = \frac{2}{3}\mathrm{e}^{-2x/3},\ x > 0$$

此时

$$\frac{f(x)}{g(x)} = \frac{3K}{2}x^{1/2}\mathrm{e}^{-x/3}$$

在上式两边求导且令其为零，则求得此比值的最大值在点

$$\frac{1}{2}x^{-1/2}\mathrm{e}^{-x/3} = \frac{1}{3}x^{1/2}\mathrm{e}^{-x/3}$$

即 $x = 3/2$ 处达到. 于是，

$$c = \max\frac{f(x)}{g(x)} = \frac{3K}{2}\left(\frac{3}{2}\right)^{1/2}\mathrm{e}^{-1/2}$$

$$= \frac{3^{3/2}}{(2\pi\mathrm{e})^{1/2}} \qquad \text{因为}K = 2/\sqrt{\pi}$$

因为

$$\frac{f(x)}{cg(x)} = (2\mathrm{e}/3)^{1/2}x^{1/2}\mathrm{e}^{-x/3}$$

故 $\Gamma(3/2,1)$ 的随机变量可以如下生成：

步骤 1：生成一个随机数 U_1，并令 $Y = -\frac{3}{2}\ln U_1$.

步骤 2：生成一个随机数 U_2.

73

步骤 3：如果 $U_2 < (2\mathrm{e}Y/3)^{1/2}\mathrm{e}^{-Y/3}$，则令 $X = Y$. 否则，转至步骤 1.

所需的平均迭代次数为

$$c = 3\left(\frac{3}{2\pi\mathrm{e}}\right)^{1/2} \approx 1.257$$

□

在上面的例子中，我们应用与伽玛分布具有同一均值的指数分布作为 $g(x)$ 的筛选法生成了伽玛随机变量. 注意到，用指数分布来生成伽玛随机变量是最有效的. 为证明这一点，假设我们想生成概率密度函数为

$$f(x) = K\mathrm{e}^{-\lambda x}x^{\alpha-1},\ x > 0$$

的随机变量，其中 $\lambda > 0, \alpha > 0$, $K = \lambda^{\alpha}/\Gamma(\alpha)$. 上述 $f(x)$ 是参数为 α 和 λ 的伽玛随机变量的密度函数，且其均值为 α/λ.

假设想利用基于参数为 μ 的指数密度函数的筛选法生成上述随机变量，因为

$$\frac{f(x)}{g(x)} = \frac{K\mathrm{e}^{-\lambda x}x^{\alpha-1}}{\mu\mathrm{e}^{-\mu x}} = \frac{K}{\mu}x^{\alpha-1}\mathrm{e}^{(\mu-\lambda)x}$$

且当 $0 < \alpha < 1$ 时

$$\lim_{x\to 0}\frac{f(x)}{g(x)} = \infty$$

故此时我们无法利用基于指数密度函数的筛选法. 因为 $\alpha = 1$ 的伽玛密度函数即为指数分布密度函数，故假设 $\alpha > 1$. 又因为当 $\mu \geqslant \lambda$ 时，

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = \infty$$

故我们限制 μ 严格小于 λ. 此时，该算法所需的平均迭代次数为

$$c(\mu) = \max_x \frac{f(x)}{g(x)} = \max_x \frac{K}{\mu}x^{\alpha-1}\mathrm{e}^{(\mu-\lambda)x}$$

为求使上式达到最大的 x 值，求导令其导数等于 0, 则有

$$0 = (\alpha-1)x^{\alpha-2}\mathrm{e}^{(\mu-\lambda)x} - (\lambda-\mu)x^{\alpha-1}\mathrm{e}^{(\mu-\lambda)x}$$

74

由此得到其最大值点为

$$x = \frac{\alpha-1}{\lambda-\mu}$$

把它反带回去，则

$$c(\mu) = \frac{K}{\mu}\left(\frac{\alpha-1}{\lambda-\mu}\right)^{\alpha-1}\mathrm{e}^{(\mu-\lambda)(\alpha-1)/(\lambda-\mu)} = \frac{K}{\mu}\left(\frac{\alpha-1}{\lambda-\mu}\right)^{\alpha-1}\mathrm{e}^{1-\alpha}$$

于是，最小化 $c(\mu)$ 的 μ 值使 $\mu(\lambda-\mu)^{\alpha-1}$ 达到最大. 其导数为

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\{\mu(\lambda-\mu)^{\alpha-1}\} = (\lambda-\mu)^{\alpha-1} - (\alpha-1)\mu(\lambda-\mu)^{\alpha-2}$$

令上式为 0，则得到最优的 μ 满足

$$\lambda-\mu=(\alpha-1)\mu$$

或

$$\mu=\lambda/\alpha$$

这就是说，用基于均值为 α/λ 的指数密度函数的筛选法生成参数为 α,λ 的伽玛随机变量所需要的迭代次数最小，且指数密度函数的均值与伽玛密度的均值相同.

下一个例子将指出如何利用筛选法生成正态随机变量.

例 5f　正态随机变量的生成　为生成标准正态随机变量 Z（即均值为 0，方差为 1 的正态），我们首先注意到其绝对值 $|Z|$ 的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2},\ 0<x<\infty \tag{5.2}$$

我们首先利用筛选法生成具有上述密度函数的随机变量，密度 $g(x)$ 采用均值为 1 的指数密度，即

75

$$g(x)=\mathrm{e}^{-x},\quad 0<x<\infty$$

此时

$$\frac{f(x)}{g(x)}=\sqrt{2/\pi}\mathrm{e}^{x-x^2/2}$$

且其最大值在使得 $x-x^2/2$ 达到最大值处取得. 由微分法可知最大值点为 $x=1$. 于是，我们取

$$c=\max\frac{f(x)}{g(x)}=\frac{f(1)}{g(1)}=\sqrt{2\mathrm{e}/\pi}$$

由于

$$\begin{aligned}\frac{f(x)}{cg(x)}&=\exp\left\{x-\frac{x^2}{2}-\frac{1}{2}\right\}\\&=\exp\left\{-\frac{(x-1)^2}{2}\right\}\end{aligned}$$

故生成 $|Z|$ 的算法如下：

步骤 1：生成参数为 1 的指数随机变量 Y.

步骤 2：生成一个随机数 U.

步骤 3：如果 $U\leqslant\exp\{-(Y-1)^2/2\}$，则令 $X=Y$. 否则，转至步骤 1.

当我们模拟得到密度函数为 (5.2) 式的随机变量 X 后（此随机变量的分布等同于标准正态随机变量的绝对值的分布），则令标准正态随机变量 Z 以等概率取 X 或 $-X$ 即可.

由于在步骤 3 中的 $U \leqslant \exp\{-(Y-1)^2/2\}$ 等价于 $-\ln U \geqslant (Y-1)^2/2$，且例 5b 证明 $-\ln U$ 是参数为 1 的指数随机变量，故上述算法等价于下面的算法：

步骤 1：生成两个独立的参数为 1 的指数随机变量 Y_1, Y_2.

步骤 2：如果 $Y_2 \geqslant (Y_1-1)^2/2$，则令 $X = Y_1$. 否则，转至步骤 1.

接受上述算法中的 Y_1 就意味着 Y_2 大于 $(Y_1-1)^2/2$，那这种大小关系大概有多少个？为回答这一问题，我们注意到 Y_2 是参数为 1 的指数随机变量，并且由指数分布的无记忆性可知，在给定它大于某值 [此值取为 Y_2 超过 $(Y_1-1)^2/2$ 的部分] 下的条件分布仍是参数为 1 的指数分布. 于是，在步骤 2 中接受 Y_1 不仅意味着我们得到了标准正态随机变量的绝对值 X，而且通过计算 $Y_2-(Y_1-1)^2/2$，我们还得到了一个参数为 1 的指数随机变量（且与 X 独立）.

76

综上所述，下述算法可同时生成一个参数 1 的指数随机变量和独立的标准正态随机变量.

步骤 1：生成参数 1 的指数随机变量 Y_1.

步骤 2：生成参数 1 的指数随机变量 Y_2.

步骤 3：如果 $Y_2-(Y_1-1)^2/2>0$，则令 $Y = Y_2-(Y_1-1)^2/2$ 且转至步骤 4. 否则，转至步骤 1.

步骤 4：生成一个随机数 U，且令

$$Z = \begin{cases} Y_1 & \text{如果 } U \leqslant \dfrac{1}{2} \\ -Y_1 & \text{如果 } U > \dfrac{1}{2} \end{cases}$$

由上一算法生成的 Z 与 Y 是独立的，且 Z 为一均值 0 方差 1 的正态随机变量，Y 是一参数为 1 的指数随机变量（如果希望生成一个均值 μ 方差 σ^2 的正态随机变量，则可取 $\mu+\sigma Z$）. □

注 1 因为 $c=\sqrt{2\mathrm{e}/\pi}\approx 1.32$，故在上一算法的步骤 2 中所需的迭代次数是一均值为 1.32 的几何随机变量.

注 2 如果我们希望生成多个标准正态随机变量，则我们可以用步骤 3 中得到的指数随机变量 Y 作为下一轮步骤 1 中所需的初始指数随机变量. 于是，平均来讲，模拟一个标准正态随机变量需要生成 $1.64(=2\times 1.32-1)$ 个指数随机变量和 1.32^2 次计算.

注 3 为确定标准正态随机变量的符号（如步骤 4），不必再生成一个新的随机数，而可以利用前面生成的第一个随机数. 也就是说，如前面用过的随机数为 $r_1, r_2, \cdots, r_k$，则可用 r_1 来确定标准正态随机变量的符号. □

当我们想模拟落在某个特定区域的随机变量时，筛选法是非常有用的. 下面这个例子将说明这一点.

例 5g 假设我们想生成取值大于 5 的 $\Gamma(2,1)$ 的随机变量. 即, 我们想生成的随机变量的概率密度函数为

$$f(x)=\frac{xe^{-x}}{\int_5^{\infty} xe^{-x}dx}=\frac{xe^{-x}}{6e^{-5}}, \quad x\geqslant 5$$

77

其中的积分是由分部积分法求得的. 由于 $\Gamma(2,1)$ 的期望为 2, 故我们利用基于取值不小于 5 的均值为 2 的指数密度函数的筛选法, 即我们取

$$g(x)=\frac{\frac{1}{2}e^{-x/2}}{e^{-5/2}},\ x\geqslant 5$$

于是,

$$\frac{f(x)}{g(x)}=\frac{e^{5/2}}{3}xe^{-x/2},\ x\geqslant 5$$

由于当 $x\geqslant 5$ 时, $xe^{-x/2}$ 是 x 的递减函数, 故此算法所需的迭代次数是均值为

$$c=\max_{x\geqslant 5}\left\{\frac{f(x)}{g(x)}\right\}=\frac{f(5)}{g(5)}=5/3$$

的几何分布随机变量. 想生成取值超过 5 的参数为 1/2 的指数随机变量, 我们应用如下事实: 由指数随机变量的无记忆性知, 此时取值超过 5 的指数随机变量仍是参数为 1/2 的指数随机变量. 即, 如果 X 是一参数为 1/2 的指数随机变量, 则 $5+X$ 与 X 具有相同的分布且超过 5. 于是, 我们有如下用来模拟具有密度函数 f 的随机变量的算法:

步骤 1: 生成一个随机数 U.

步骤 2: 令 $Y=5-2\log(U)$.

步骤 3: 生成一个随机数 U.

步骤 4: 如果 $U\leqslant \frac{e^{5/2}}{5}Ye^{-Y/2}$, 则令 $X=Y$ 且停止; 否则回到步骤 1. □

正如在例 5f 中我们利用基于指数密度函数的筛选法模拟正态随机变量一样, 我们也可以利用基于指数密度函数的筛选法很有效地模拟落在某个特定区间内的正态随机变量. 细节 (包括指数分布最佳均值的确定) 将在 8.8 节给出.

5.3 生成正态随机变量的极坐标法

设 X 和 Y 是两个独立的标准正态随机变量, 以 R 和 θ 记向量 (X,Y) 的极坐标. 则有 (见图 5-2)

78

$$R^2=X^2+Y^2$$
$$\tan\theta=\frac{Y}{X}$$

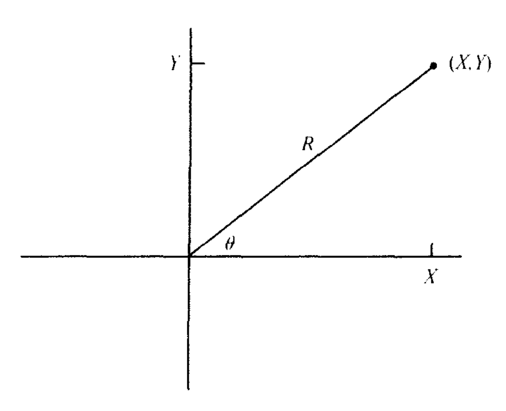


图 5-2 点 (X,Y) 的极坐标表示

因为 X 与 Y 是独立的，故它们的联合密度为二者密度的乘积，即

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-y^2/2} \\ &= \frac{1}{2\pi}e^{-(x^2+y^2)/2} \end{aligned} \tag{5.3}$$

为确定 R^2 和 Θ 的联合密度 $f(d,\theta)$，我们做下面变量替换：

$$d = x^2 + y^2,\ \theta = \arctan\left(\frac{y}{x}\right)$$

容易验证此变换的雅可比式，即上述变换关于 x 和 y 的偏导数矩阵的行列式等于 2. 于是，由等式 (5.3) 知，R^2 和 Θ 的联合密度为

$$f(d,\theta) = \frac{1}{2}\frac{1}{2\pi}e^{-d/2},\ 0 < d < \infty,\ 0 < \theta < 2\pi$$

由于上式等于均值 2 的指数密度 $(\frac{1}{2}e^{-d/2})$ 和 $(0,2\pi)$ 上均匀密度 $[(2\pi)^{-1}]$ 的乘积，故有

$$R^2\text{与}\Theta\text{独立，且}R^2\text{是均值 2 的指数随机变量，}\Theta\text{是}(0,2\pi)\text{上的均匀随机变量.} \tag{5.4}$$

79

于是，我们可以先用等式 (5.4) 生成极坐标，之后把它们变成直角坐标就得到了两个独立的标准正态随机变量. 这即是下面的算法：

步骤 1：生成随机数 U_1, U_2.

步骤 2：$R^2 = -2\ln U_1$(R^2 是均值 2 的指数随机变量)，$\Theta = 2\pi U_2$[Θ 是 $(0,2\pi)$ 上均匀随机变量].

步骤 3：令

$$X = R\cos\Theta = \sqrt{-2\ln U_1}\cos(2\pi U_2),\quad Y = R\sin\Theta = \sqrt{-2\ln U_1}\sin(2\pi U_2) \tag{5.5}$$

等式 (5.5) 的变换被称为Box-Muller 变换.

然而，利用 (5.5) 式的 Box-Muller 变换生成两个独立标准正态的计算效率不高，其原因是它需要计算正余弦两个三角函数. 但是，我们可以通过直接计算某随机角度的正余弦 (而不是对于某随机数 U，直接计算 $2\pi U$ 的正余弦) 来克服其费时的不足. 实际上，我们注意到，如果 U 是 (0,1) 上的随机数，则 $2U$ 是 (0,2) 上的随机数，$2U-1$ 是 $(-1,1)$ 上的随机数. 于是，如果 U_1,U_2 是两个随机数，且令

$$V_1 = 2U_1 - 1$$
$$V_2 = 2U_2 - 1$$

则 (V_1,V_2) 在中心为原点、面积为 4 的正方形中均匀分布 (见图 5-3) .

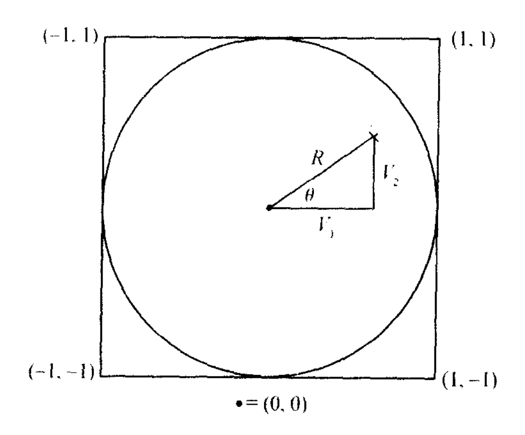


图 5-3　随机数正余弦函数的计算

现假设我们连续产生随机数对 (V_1,V_2)，直至得到一对落入中心在原点半径为 1 的圆内，即满足 $V_1^2+V_2^2\leqslant 1$. 注意到满足此条件的 (V_1,V_2) 在此圆内均匀分布. 如以 R 和 Θ 记此点的极坐标，则不难验证，R 与 Θ 是独立的，且 R^2 在 (0,1) 上均匀分布 (见习题 21)，Θ 在 $(0,2\pi)$ 上均匀分布. 因为此角度 Θ 是随机的，故由上述可知，一个随机角度 Θ 的正余弦函数可以如下计算：

$$\sin\Theta = \frac{V_2}{R} = \frac{V_2}{\sqrt{(V_1^2+V_2^2)^{1/2}}}$$

$$\cos\Theta = \frac{V_1}{R} = \frac{V_1}{\sqrt{(V_1^2+V_2^2)^{1/2}}}$$

80

其中 (V_1,V_2) 为满足条件的随机点. 于是，综合 Box-Muller 变换 (5.5) 和上式，一对独立的标准正态随机变量可以如下生成：生成一个随机数 U 后，令

$$X = (-2\ln U)^{1/2}\frac{V_1}{(V_1^2+V_2^2)^{1/2}}$$
$$Y = (-2\ln U)^{1/2}\frac{V_2}{(V_1^2+V_2^2)^{1/2}} \tag{5.6}$$

另外，由于 $R^2 = V_1^2 + V_2^2$ 在 (0,1) 上均匀分布且它与随机角度 Θ 独立，故它可以替代 (5.6) 式中的随机数 U. 于是，如令 $S = R^2$，则我们知

$$X = (-2\ln S)^{1/2}\frac{V_1}{S^{1/2}} = V_1\left(\frac{-2\ln S}{S}\right)^{1/2}$$

$$Y = (-2\ln S)^{1/2}\frac{V_2}{S^{1/2}} = V_2\left(\frac{-2\ln S}{S}\right)^{1/2}$$

是独立的标准正态随机变量，其中 (V_1, V_2) 是中心在原点半径为 1 的圆内的随机点，且 $S = V_1^2 + V_2^2$.

综上所述，我们有如下的产生一对独立的标准正态随机变量的极坐标算法：

步骤 1：生成随机数 U_1, U_2.

步骤 2：令 $V_1 = 2U_1 - 1, V_2 = 2U_2 - 1, S = V_1^2 + V_2^2$.

步骤 3：如果 $S > 1$，则转至步骤 1.

步骤 4：产生独立的标准正态随机变量：

81

$$X = \sqrt{\frac{-2\ln S}{S}}V_1,\quad Y = \sqrt{\frac{-2\ln S}{S}}V_2$$

由于正方形内的随机点落入圆内的概率为 $\pi/4$(圆与正方形面积之比)，故极坐标法中步骤 1 所需的平均迭代次数为 $4/\pi = 1.273$. 于是，为生成两个独立的标准正态随机变量，平均来讲，共需 2.546 个随机数、一次对数、一次平方根、一次除法和 4.546 次乘法.

5.4 泊松过程的生成

如果我们希望生成一个参数 λ 的泊松过程的前 n 个事件发生的时间，则我们可以利用如下结论：此过程两个相继事件发生的间隔时间为参数 λ 的指数随机变量. 于是，生成此过程的一种方法就是先生成这些间隔时间. 如以 $U_1, U_2, \cdots, U_n$ 记 n 个随机数，且令 $X_i = -\frac{1}{\lambda}\ln U_i$，则 X_i 即可作为此泊松过程第 $i-1$ 个和第 i 个事件发生的间隔时间. 由于第 j 个事件发生的时间等于前 j 个间隔时间之和，则前 n 个事件发生的时间分别为 $\sum_{i=1}^{j} X_i$，$j = 1, \cdots, n$.

如果我们想模拟此泊松过程时刻 T 前的状态，则我们可以利用上述方法相继生成其间隔时间，直至其和超过 T 为止. 也就是说，我们可以利用下述算法模拟一个参数为 λ 的泊松过程在 $(0,T)$ 时间段内的事件发生时间，其中 t 表示时间，I 表示时间 t 内发生的事件数，$S(I)$ 表示最近事件发生的时间.

一个参数为 λ 的泊松过程在时刻 T 前的状态模拟

步骤 1：$t=0, I=0$.

步骤 2：生成一个随机数 U.

步骤 3：$t=t-\frac{1}{\lambda}\ln U$. 如果 $t>T$，停止.

步骤 4：$I=I+1, S(I)=t$.

步骤 5：转至步骤 2.

上述算法中 I 的最终值表示时间 T 内发生的事件数，$S(1),\cdots,S(I)$ 递增且表示 I 个事件发生的时间.

另一种模拟参数为 λ 的泊松过程时刻 T 前状态的方法是先模拟时刻 T 发生的事件总数 $N(T)$. 由于 $N(T)$ 是一均值为 λT 的泊松随机变量，这可以利用第 4 章所给的方法模拟它. 如果 $N(T)$ 的模拟值为 n，则生成 n 个随机数 $U_1,\cdots,U_n$，且取 $\{TU_1,\cdots,TU_n\}$ 为时刻 T 时此泊松过程的事件时间集.

82

为验证上述方法的正确性，以 $N(t)$ 记集合 $\{TU_1,\cdots,TU_n\}$ 中小于 t 的元素个数，则我们必须说明 $N(t), 0\leqslant t\leqslant T$ 是一个泊松过程. 为证明它是一个具有独立增量的过程，记 $I_1,\cdots,I_r$ 为区间 $[0,T]$ 中的 r 个不相交的时间区间. 如果 TU_i 属于区间 $I_i(i=1,\cdots,r)$，则称第 i 个泊松事件为类型 i. 如果 TU_i 不属于 r 区间的任何一个，则称第 i 个泊松事件为类型 $r+1$. 如 p_i，$i\leqslant r$ 表示区间 I_i 的长度与 T 之商，$p_{r+1}=1-\sum_{i=1}^{r}p_i$，由于 U_i，$i\geqslant 1$ 独立，则 $N(T)$ 个泊松事件中的每一个都可以独立地以概率 p_i 被归类到类型 $i, i=1,\cdots,r+1$. 于是，由 2.8 节的结论知，落在 r 个不相交区间的事件数 $N_1,\cdots,N_r$ 是独立的泊松随机变量，且期望 $E[N_i]$ 等于 λ 与区间 I_i 的长度之积. 由此可知，$N(T), 0\leqslant t\leqslant T$ 是一具有独立增量的平稳过程. 由于落在任一长度为 h 的区间内的事件数服从均值为 λh 的泊松分布，故我们有

$$\lim_{h\to 0}\frac{P\{N(h)=1\}}{h}=\lim_{h\to 0}\frac{\lambda h\mathrm{e}^{-\lambda h}}{h}=\lambda$$

和

$$\lim_{h\to 0}\frac{P\{N(h)\geqslant 2\}}{h}=\lim_{h\to 0}\frac{1-\mathrm{e}^{-\lambda h}-\lambda h\mathrm{e}^{-\lambda h}}{h}=0$$

由此结论得证.

如果我们仅想模拟一个泊松随机变量过程事件发生时间的集合，则上述方法比利用其间隔时间服从指数分布的模拟方法更有效. 然而，我们一般均希望事件发生时间是递增排序的，故我们还要将 $TU_i, i=1,\cdots,n$ 排序.

5.5 非齐次泊松过程的产生

如放松泊松过程中的平稳增量条件，则得到非齐次泊松过程. 非齐次泊松过程是建模中常用计数过程中非常重要的一种. 从概率的角度看，非齐次泊松过程的参数，即单位时间内事件出现的个数，并不需要是常数，而可以随时间变化. 一般来说，对于一个用非齐次泊松过程来刻划的数学模型，我们很难得到它的解析解. 于是，非齐次泊松过程的应用并不非常广泛. 然而，由于我们可以用模拟方法分析这样的模型，故我们期望这样的数学模型会得到广泛的应用.

83

假设我们希望模拟强度函数为 $\lambda(t)$ 的非齐次泊松过程时刻 T 前的状态. 被称为*稀松*(thinning) 或*随机抽样法*是我们给出的第一种方法. 在选取满足

$$\lambda(t)\leqslant\lambda,\ \forall\, t\leqslant T$$

的 λ 之后，如在第 2 章证明的那样，我们可以通过随机选取参数为 λ 的泊松过程的事件发生时间来生成非齐次泊松过程. 也就是说，如果一个参数为 λ 的泊松过程在时刻 t 发生的事件以概率 $\lambda(t)/\lambda$ 被记录，则被记录到的事件列就构成一个强度为 $\lambda(t)$, $0\leqslant t\leqslant T$ 的非齐次泊松过程. 于是，我们可以通过对泊松过程的事件进行随机计数来产生所需的非齐次泊松过程，其算法如下：

在时刻 T 前一个泊松过程的产生

步骤 1：$t=0, I=0$.

步骤 2：生成一个随机数 U.

步骤 3：$t=t-\frac{1}{\lambda}\ln U$. 如果 $t>T$，则停止.

步骤 4：生成一个随机数 U.

步骤 5：如果 $U\leqslant\lambda(t)\lambda$，则令 $I=I+1, S(I)=t$.

步骤 6：转至步骤 2.

在上述算法中，$\lambda(t)$ 是强度函数，λ 满足 $\lambda(t)\leqslant\lambda$，$I$ 的最终值表示时刻 T 前发生的事件数，$S(1),\cdots,S(I)$ 是事件发生的时刻.

上述算法之所以被称为稀松法，是由于它仅取齐次泊松过程的部分事件. 如以被记录到的事件数作为有效与否的标准，则当 $\lambda(t)$ 与 λ 在整个区间上相差不多时，上述算法是非常有效的. 于是，若把整个区间划分成若干个子区间，然后再在子区间上应用上述算法，则可以改进上述算法的效率. 现选取适当的 k, t_i 和 λ_i，使得

$0 = t_0 < t_1 < t_2 < \cdots < t_k < t_{k+1} = T$, 且

$$\lambda(s) \leqslant \lambda_i, \text{ 如果} t_{i-1} \leqslant s < t_i, \ i = 1, \cdots, k+1 \tag{5.7}$$

84 为生成区间 (t_{i-1}, t_i) 上的非齐次泊松过程，我们先生成参数为 λ_i 的指数随机变量，然后以概率 $\lambda(s)/\lambda_i$ 接受在 $s \in (t_{i-1}, t_i)$ 时刻发生的事件. 由指数分布的无记忆性和其参数的数乘性可知，从一个子区间到另一个子区间，其效率并没有损失. 也就是说，如果现在我们考虑的区间为 (t_{i-1}, t_i) 且以 X 记参数为 λ_i 的指数分布随机数，即 $t + X > t_i$，则 $\lambda_i[X - (t_i - t)]/\lambda_{i+1}$ 可作为下一区间的参数为 λ_{i+1} 的指数分布随机数.

对于强度 $\lambda(s)$ 满足关系式 (5.7) 的非齐次泊松过程，我们可用下述算法来生成它的随机数. 在下面算法中，t 表示当前时刻，J 表示区间 (如 $t_{j-1} \leqslant t < t_j$，则 $J = j$), I 表示前面发生的事件数, $S(1), \cdots, S(I)$ 表示事件发生的时间.

在时刻 T 前一个非齐次泊松过程的产生

步骤 1：$t = 0, J = 1, I = 0$.

步骤 2：生成一个随机数 U 且令 $X = -\frac{1}{\lambda_J} \ln U$.

步骤 3：如果 $t + X > t_J$，则转至步骤 8.

步骤 4：$t = t + X$.

步骤 5：生成一个随机数 U.

步骤 6：如果 $U \leqslant \lambda(t)/\lambda_J$，则 $I = I + 1, S(I) = t$.

步骤 7：转至步骤 2.

步骤 8：如果 $J = k + 1$，则停止.

步骤 9：$X = (X - t_J + t)\lambda_J/\lambda_{J+1}, t = t_J, J = J + 1$.

步骤 10：转至步骤 3.

对于区间 (t_{i-1}, t_i)，记

$$\lambda_i \equiv \inf\{\lambda(s) : t_{i-1} \leqslant s < t_i\}$$

且假设 $\lambda_i > 0$. 此时，我们将不直接应用稀松算法，而是先在第 i 个区间产生参数为 λ_i 的泊松过程，之后再模拟强度为 $\lambda(s) = \lambda(s) - \lambda_i[s \in (t_{i-1}, t_i)]$ 的非齐次泊松过程. (为模拟此泊松过程而生成的最后一个指数随机变量，虽然它超出了要求区间，但也不必浪费掉，因为它经适当变换后还可重新再用.) 把上面产生的两个过程叠加起来，则得到了此区间所需的过程. 我们这样做的原因是它将节省产生均值为 $\lambda_i(t_i - t_{i-1})$ 的泊松分布随机数所需的均匀分布随机变量. 例如，我们考虑

85

$$\lambda(s) = 10 + s, \quad 0 < s < 1$$

利用 $\lambda=11$ 的稀松法，可能生成其期望数为 11 的事件列，其中每一个事件都需用一个随机数来确定此事件是否被接受. 换言之，先产生参数为 10 的泊松过程与参数为 $\lambda(s)=s,\ 0<s<1$ 的非齐次泊松过程（利用参数为 $\lambda=1$ 的稀松法），把二者合并后得到的事件数与所要求的同分布，但需验证被接受事件数的期望等于 11.

产生强度为 $\lambda(t),\ t>0$ 的非齐次泊松过程的另一个方法是直接产生相继事件的发生时间. 以 $S_1,S_2,\cdots$ 记此过程相继事件的发生时间. 由于这些随机变量显然是相关的，故我们在生成它们时是先生成 S_1，然后再利用 S_1 的生成值去模拟 S_2，依此类推.

我们注意到，如一个事件在时间 s 发生，故它独立于时刻 s 前出现的事件，且到下一个事件发生的时间间隔的分布 F_s 为

$$\begin{aligned}
F_s(x)&=P\{\text{从}s\text{到下一事件的时间间隔小于}x|\text{在时刻}s\text{发生一事件}\}\\
&=P\{\text{下一事件在时刻}x+s\text{前发生}|\text{在时刻}s\text{发生一事件}\}\\
&=P\{\text{下一事件在时间}s\text{和}s+x\text{间发生}|\text{在时刻}s\text{发生一事件}\}\\
&=P\{\text{下一事件在时间}s\text{和}s+x\text{间发生}\}\ \text{由独立增量}\\
&=1-P\{\text{在时间}s\text{和}s+x\text{间没有事件发生}\}\\
&=1-\exp\left(-\int_s^{s+x}\lambda(y)\mathrm{d}y\right)\\
&=1-\exp\left(-\int_0^{x}\lambda(s+y)\mathrm{d}y\right)
\end{aligned}\tag{5.8}$$

于是，我们可以由分布 F_0 产生 S_1，如记其值为 s_1，则 S_2 的模拟值为 s_1 加上由分布 F_{s_1} 产生的随机数，如记此和为 s_2，则 S_3 的模拟值为 s_2 加上由分布 F_{s_2} 产生的随机数，依此类推，就得到了事件发生时间 $S_1,S_2,\cdots$ 的模拟值. 上述由分布进行随机模拟的方法肯定依赖其分布形式. 在下面的例子中，分布 F_s 的逆函数容易求得，故我们可以利用逆变换法.

例 5h 设 $\lambda(t)=1/(t+a),\ t\geqslant 0$，其中 a 为正常数. 则

$$\int_0^x\lambda(s+y)\mathrm{d}y=\int_0^x\frac{1}{s+y+a}\mathrm{d}y=\ln\left(\frac{x+s+a}{s+a}\right)$$

86

于是，由等式 (5.8) 有

$$F_s(x)=1-\frac{s+a}{x+s+a}=\frac{x}{x+s+a}$$

为求其逆，设 $x=F_s^{-1}(u)$，则

$$u=F_s(x)=\frac{x}{x+s+a}$$

或等价地有

$$x = \frac{u(s+a)}{1-u}$$

即

$$F_s^{-1}(u) = (s+a)\frac{u}{1-u}$$

因此，相继事件的发生时间 $S_1, S_2, \cdots$ 可如下产生：生成随机数 $U_1, U_2, \cdots$ 后，依次取

$$S_1 = \frac{aU_1}{1-U_1}$$

$$S_2 = S_1 + (S_1 + a)\frac{U_2}{1-U_2} = \frac{S_1 + aU_2}{1-U_2}$$

一般地有

$$S_j = S_{j-1} + (S_{j-1} + a)\frac{U_j}{1-U_j} = \frac{S_{j-1} + aU_j}{1-U_j}, \quad j \geqslant 2$$

□

习　　题

1. 给出具有下述生成密度函数的随机变量的方法.

$$f(x) = \mathrm{e}^x/(\mathrm{e}-1),\ 0 \leqslant x \leqslant 1$$

2. 给出具有如下生成密度函数的随机变量的方法.

87

$$f(x) = \begin{cases} \dfrac{x-2}{2} & \text{如果} 2 \leqslant x \leqslant 3 \\ \dfrac{2-x/3}{2} & \text{如果} 3 \leqslant x \leqslant 6 \end{cases}$$

3. 应用逆变换法生成如下分布函数的随机变量.

$$F(x) = \frac{x^2+x}{2},\ 0 \leqslant x \leqslant 1$$

4. 给出生成分布函数为

$$F(x) = 1 - \exp(-\alpha x^\beta),\ 0 < x < \infty$$

的随机变量方法. (具有上述分布函数的随机变量称为威布尔随机变量.)

5. 给出具有如下生成密度函数的随机变量的方法.

$$f(x) = \begin{cases} e^{2x} & -\infty < x < 0 \\ e^{-2x} & 0 < x < \infty \end{cases}$$

6. 设 X 为均值 1 的指数随机变量. 给出模拟分布为给定 $X < 0.05$ 下 X 的条件分布的随机变量的有效算法, 其密度为

$$f(x) = \frac{e^{-x}}{1 - e^{-0.05}},\ 0 < x < 0.05$$

生成 1000 个这样的随机变量, 并用它们估计 $E[X|X < 0.05]$. 另外, 求 $E[X|X < 0.05]$ 的精确值.

7. (复合法) 设产生分布为 F_i, $i = 1, 2, \cdots, n$ 的随机变量相对容易. 我们如何产生分布函数为

$$F(x) = \sum_{i=1}^{n} p_i F_i(x)$$

的随机变量? 其中 p_i, $i = 1, \cdots, n$ 是和为 1 的非负数.

8. 利用习题 7 的结论, 给出生成具有如下分布函数的随机变量的算法:
   (a) $F(x) = \frac{x+x^3+x^5}{3}$, $0 \leqslant x \leqslant 1$.
   (b) $F(x) = \begin{cases} \frac{1-e^{-2x}+2x}{3} & 如果 0 < x < 1 \\ \frac{3-e^{-2x}}{3} & 如果 1 < x < \infty \end{cases}$
   (c) $F(x) = \sum_{i=1}^{n} \alpha_i x^i$, $0 \leqslant x \leqslant 1$, 其中 $\alpha_i \geqslant 0, \sum_{i=1}^{n} \alpha_i = 1$.

88

9. 给出生成分布函数为

$$F(x) = \int_0^\infty x^y e^{-y} dy,\ 0 \leqslant x \leqslant 1$$

的随机变量的算法. [提示: 考虑利用习题 7 中的复合法. 特别地, 以 F 记 X 的分布函数, 且假设给定 $Y = y$ 下 X 的条件分布为

$$P\{X \leqslant x | Y = y\} = x^y,\ 0 \leqslant x \leqslant 1]$$

10. 设一个灾害保险公司有 1000 个客户, 且每个客户在下月提出索赔的概率为 0.05, 并相互独立, 其索赔金额是独立的均值为 800 美元的指数随机变量. 请用模拟方法估计其索赔金额之和超过 5 万美元的概率.

11. 给出一种生成 3 个指数随机变量的算法, 并与例 5c 后的成对生成随机变量的算法进行计算量的比较.

12. 假设容易生成分布为 F_i, $i = 1, \cdots, n$ 的随机变量, 我们如何生成来自如下分布的随机变量?

(a) $F(x)=\prod_{i=1}^{n} F_i(x)$.

(b) $F(x)=1-\prod_{i=1}^{n}[1-F_i(x)]$.

[提示：如果 $X_i, i=1,\cdots,n$ 独立，且 X_i 的分布为 F_i，则什么样的随机变量的分布为 F？]

13. 利用筛选法和习题 12 的结果，给出两种有别于逆变换法的方法生成具有如下分布函数的随机变量.

$$F(x)=x^n,\ 0\leqslant x\leqslant 1$$

14. 设 G 为一分布函数，g 为其密度. 对于常数 $a<b$，我们希望生成具有如下分布函数的随机变量.

$$F(x)=\frac{G(x)-G(a)}{G(b)-G(a)},\ a\leqslant x\leqslant b$$

89

(a) 如果 X 的分布函数为 G，则 F 是给定什么条件下 X 的条件分布？

(b) 证明此时的筛选法为：生成分布函数为 G 的随机变量 X 后，如果它落在 a 与 b 之间，则接受它.

15. 给出生成密度函数为

$$f(x)=x\mathrm{e}^{-x},\ 0\leqslant x<\infty$$

的随机变量的两种方法，并比较二者的效率.

16. 设随机变量 X 的密度函数为

$$f(x)=\frac{1}{2}x^2\mathrm{e}^{-x},\ x>0$$

现用带有参数为 λ 的指数密度的筛选法生成此模拟值. 并求最优的 λ 值使此算法的平均迭代次数最少.

17. 给出生成密度函数为

$$f(x)=30(x^2-2x^3+x^4),\ 0\leqslant x\leqslant 1$$

的随机变量的方法，并讨论此方法的效率.

18. 给出生成密度函数为

$$f(x)=\frac{1}{0.000\,336}x(1-x)^3,\ 0.8<x<1$$

的随机变量 X 的有效方法.

19. 在例 5f 中，我们用带有参数为 1 的指数密度函数的筛选法给出了生成一个正态随机变量的方法. 证明在所有的指数密度函数 $g(x)=\lambda\mathrm{e}^{-\lambda x}$ 中，$\lambda=1$ 使上述算法的迭代次数最少.

20. 编写用例 5f 中的方法生成正态随机变量的程序.
21. 设 (X,Y) 在半径为 1 的圆上均匀分布, 证明: 如果 R 是点 (X,Y) 到圆心的距离, 则 R^2 在 (0,1) 上均匀分布.
22. 编写生成参数为 λ 的泊松过程在时刻 T 前状态的程序.
23. 某项工作的完成必须依次经历 k 道工序, 一名工人完成第 i 道工序所需的时间是一个参数为 λ_i 的指数随机变量, $i=1,\cdots,k$. 然而, 一名工人在完成第 i 道工序后接着从事下一道工序的概率仅为 α_i, $i=1,2,\cdots,k-1$, 即一名工人在完成第 i 道工序后将以概率 $1-\alpha_i$ 不再从事这项工作. 如以 X 记一名工人从事此项工作的时间总量, 则 X 被称为Cox 随机变量. 编写模拟此随机变量的程序.

90

24. 设公共汽车到达某项比赛场所的时间服从参数 5(即每小时有 5 个事件发生) 的泊松过程, 每辆公共汽车均等可能地有 $20,21,\cdots,40$ 个球迷乘座, 且不同公共汽车上的球迷数是相互独立的. 编写模拟 $t=1$ 时刻抵达此比赛场所的球迷数的程序.
25. (a) 利用稀松法编写模拟前 10 个时间单元内强度为

$$\lambda(t)=3+\frac{4}{t+1}$$

的非齐次泊松过程的程序.

(b) 给出一种改进本例稀松法的方法.

26. 给出有效模拟前 10 个时间单元内强度为

$$\lambda(t)=\begin{cases}\dfrac{t}{5} & 0<t<5\\ 1+5(t-5) & 5<t<10\end{cases}$$

的非齐次泊松过程的算法.

参考文献

[1] Dagpunar, T., *Principles of Random Variate Generation.* Clarendon Press, Oxford, 1988.

[2] Devroye, L., *Nonuniform Random Variate Generation.* Springer-Verlag, New York, 1986.

[3] Fishman, G. S., *Principles of Discrete Event Simulation.* Wiley, New York, 1978.

[4] Knuth, D., *The Art of Computer Programming*, Vol.2, 2nd ed., *Seminumerical Algorithms*, Addison-Wesley, Reading, MA, 2000.

[5] Law, A. M., and W. D. Kelton, *Simulation Modelling and Analysis*, 3rd ed. McGraw-Hill, New York, 1997.

[6] Lewis, P. A. W., and G. S. Shedler, *Simulation of Nonhomogeneous Poisson Processes by Thinning, Nav. Res. Log. Quart.*, **26**, 403–413, 1979.

[7] Marsaglia, G., *Generating Discrete Random Variables in a Computer, Commun. Assoc. Comput. Mach.*, **6**, 37-38, 1963.

[8] Morgan, B. J. T., *Elements of Simulation.* Chapman and Hall, London, 1983.

[9] Ripley, B. D., *Computer Generation of Random Variables:A Tutorial, Inst. Statist. Rev.*, **51**, 301–319, 1983.

[10] Ripley, B. D., *Stochastic Simulation.* Wiley, New York, 1986.

[11] Rubenstein, R. Y., *Simulation and the Monte Carlo Method.* Wiley, New York, 1981.

[12] Schmeiser, B. W., *Random Variate Generation, a Survey, Proc. 1980 Winter Simulation Conf.*, Orlando, FL; pp.79-104, 1980.

91

第 6 章　离散事件模拟法

引　言

一个概率模型的模拟包括产生此模型的随机机理和观察此模型相应结果两部分. 要模拟某一个模型，我们就必须确定某些感兴趣的量. 然而，因为模型随时间在变化且其机理的逻辑结构很复杂，故人们一般不能明确地了解其如何变化以及如何确定上述感兴趣的量. 基于"离散事件"的思想，一个通用的框架可以帮助人们模拟一个随机时间变化的模型及确定其中某些感兴趣的量. 基于这个框架的模拟方法就被称为*离散事件模拟法*.

6.1　离散事件模拟法

变量和事件是离散事件模拟法中最重要的元素. 在模拟中，我们始终紧盯某些变量. 一般地，有如下三种常用的变量类型：时间变量、计数变量和系统状态变量.

变量

1. 时间变量 t	表示模拟所用的时间总量
2. 计数变量	这些变量表示时刻 t 某些事件出现的次数
3. 系统状态变量 (SS)	此变量描述系统在时刻 t 的状态

只要出现一个事件，上述变量的值就要改变或更新，我们就要收集相应感兴趣的数据作为输出. 为确定下一事件何时出现，我们需要一个事件列表 (此列表给出后面最近的事件和这些事件出现的时间表). 只要出现一个事件，我们就要重新设置时间变量、状态变量、计数变量和收集到的相应数据. 这样做的目的在于我们可以及时地追踪随时间而变化的系统.

由于上面仅给出了离散事件模拟法的大致思想，故下面几个例子对于了解如何应用此种方法是很有帮助的. 在 6.2 节，我们考虑单服务员等候或排队系统的模拟. 6.3 节和 6.4 节考虑多服务员排队系统的模拟，其中 6.3 节假设各服务员是串联排列的，而 6.4 节则假设各服务员是并联排列的. 6.5 节是一仓储模型，6.6 节是一保险风险模型，6.7 节是一个机器维修模型. 在 6.8 节我们考虑一个股票期权模型.

在所有排队模型中，我们假设顾客到达时间服从一个带有有界强度函数 $\lambda(t)$, $t>0$ 的非齐次泊松过程. 在模拟这些模型中，下面子程序将被用来生成随机变量

T_s，其中 T_s 表示时间 s 后第一位顾客的到来时间.

对于给定的 $\lambda(t), t > 0$，假设 λ 满足 $\lambda(t) \leqslant \lambda, \forall\, t$. 则生成 T_s 的子程序为：

生成 T_s 的子程序

步骤 1：设 $t = s$.

步骤 2：生成随机数 U.

步骤 3：记 $t = t - \frac{1}{\lambda}\ln U$.

步骤 4：生成随机数 U.

步骤 5：如果 $U \leqslant \lambda(t)/\lambda$，记 $T_s = t$ 并停止.

步骤 6：转至步骤 2.

6.2　单服务员排队系统

在一个服务站，假设顾客到达时间服从一个强度函数为 $\lambda(t), t \geqslant 0$ 的非齐次泊松过程，且仅有一个服务员. 当服务员空闲时，到达的顾客可以得到即时的服务；而当服务员工作时，新来的顾客要排队等候服务. 另外，我们还假设服务员完成一位顾客的服务后，如果有顾客排队等候服务，则他转而服务下一个等候时间最长的顾客（这种服务称之为"先来先得"）；如果没有顾客排队，则他空闲下来等候下一位顾客的到来. 假设每一位顾客所需的服务时间是一概率分布为 G 的随机变量，且独立于其他顾客的服务时间和到达时间. 另外，假设时间 T 后下班，不再允许顾客进入服务站等候服务，即使服务员已经完成了对所有在时间 T 前进入服务站的顾客的服务，其中 T 是一固定值.

94

在上面系统中，假设我们感兴趣且想确定的量为：(a) 顾客在系统中所花的平均时间；(b) 时间 T 后，最后一位顾客离开服务站的平均时间，即服务员可以回家时的平均加班时间.

在模拟上述系统时，我们用到的变量如下：

时间变量　　t

计数变量　　N_A：时刻 t 到达的顾客数

　　　　　　　N_D：时刻 t 离开的顾客数

系统状态变量　　n：时刻 t 时服务站中的顾客数

因为顾客的到来或离开均可改变上述变量，故我们称它们为"事件". 也就是说，此系统中有两类事件：离开和到达. 于是，事件列表包含下一顾客的到达时间和正在接受服务顾客的离开时间，即事件列表为

$$\mathbf{EL} = t_A, t_D$$

其中 t_A 是时间 t 后下一顾客的到达时间，t_D 是正在接受服务的顾客的离开时间. 如果此时没有顾客接受服务，则取 t_D 为 ∞.

我们收集的输出变量为：第 i 个顾客的到达时间 $A(i)$ 和离开时间 $D(i)$; 时间 T 后最后一位顾客的离开时间 T_p.

开始模拟之前，变量和事件的初始值取为：

初始化

取 $t = N_A = N_D = 0$.

取 $\mathbf{SS} = 0$.

生成 T_0，并取 $t_A = T_0, t_D = \infty$.

为更新此系统，我们就要改变时间直至下一个事件出现. 至于出现哪个事件，则要考虑多种情形，且依赖于事件列表中的较小者. 下面以 Y 记分布为 G 的服务时间随机变量.

95

$$t = \text{时间变量}, \quad \mathbf{SS} = n, \quad \mathbf{EL} = t_A, t_D$$

情形 1：$t_A \leqslant t_D, t_A \leqslant T$

重新取 $t = t_A$(转到时刻 t_A).

重新取 $N_A = N_A + 1$(因为在时刻 t_A 多了一位顾客).

重新取 $n = n + 1$(因为此时多了一位顾客).

生成 T_t，并重新取 $t_A = T_t$(这是下一位顾客的到达时间).

如果 $n = 1$，则生成 Y 并重新取 $t_D = t + Y$(因为此时服务站无顾客，故必须产生下一位顾客所需的服务时间).

收集输出数据 $A(N_A) = t$(因为顾客 N_A 在时间 t 到达).

情形 2：$t_D < t_A, t_D \leqslant T$

重新取 $t = t_D$.

重新取 $n = n - 1$.

重新取 $N_D = N_D + 1$(因为一人在时间 t 离开).

如果 $n = 0$，则重新取 $t_D = \infty$; 否则，生成 Y 并重新取 $t_D = t + Y$.

收集输出数据 $D(N_D) = t$(因为顾客 N_D 刚离开).

情形 3：$\min(t_A, t_D) > T, n > 0$

重新取 $t = t_D$.

重新取 $n = n - 1$.

重新取 $N_D = N_D + 1$.

如果 $n > 0$，则生成 Y 并重新取 $t_D = t + Y$.

收集输出数据 $D(N_D) = t$.

情形 4：$\min(t_A, t_D) > T, n = 0$

收集输出数据 $T_p = \max(t - T, 0)$.

上述算法的流程图见图 6-1. 每到“停止”块时，我们就收集数据 N_A(到达顾客的总数，它等于离开顾客的总数 N_D). 对每个顾客 $i, i = 1, \cdots, N_A$，我们有其到达和离开时间 $A(i)$ 和 $D(i)$，于是 $D(i) - A(i)$ 就表示顾客 i 在系统中的时间. 最后，我们可以得到最后一位顾客离开的时间 T_p. 每收集一次数据，我们就说完成了一次模拟. 每次模拟之后，我们重新对变量初始化并进行另一次模拟，直至我们认为已收集到足够多的数据（何时结束模拟的问题将在第 7 章考虑）. 收集到的所有 T_p 的平均即为最后一位顾客离开的平均时间；所有观察到的 $D - A$ 的平均即是一位顾客在系统中所花时间的估计.

96

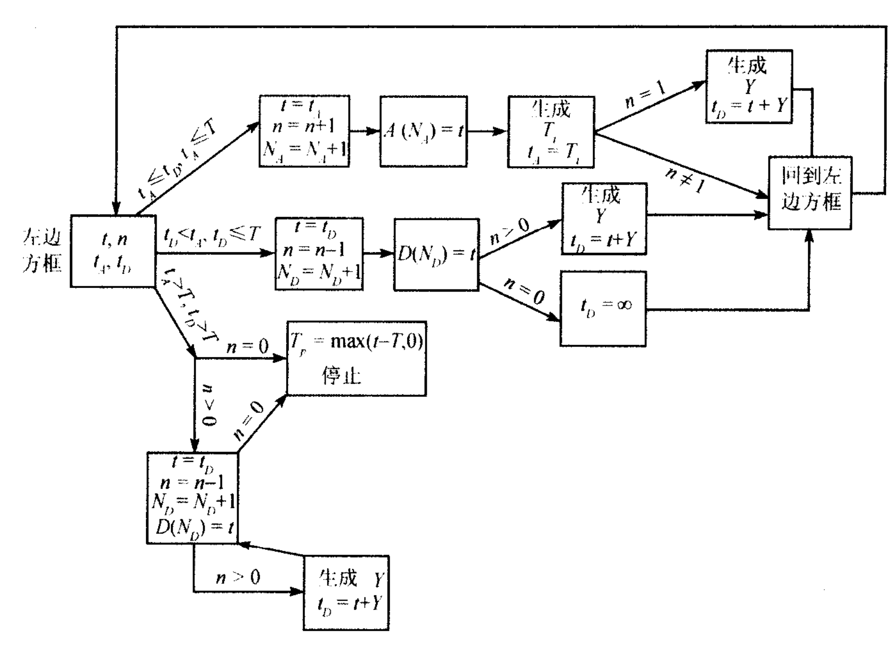


图 6-1　模拟单服务员系统的流程图

注　如果我们想存储输出数据以得到即时的顾客人数，则无论是哪个事件出现，都必须输出系统状态和时间变量对 (n, t). 例如，如果输出数据为 $(1, 4)$ 和 $(0, 6)$，且以 $n(t)$ 记时刻 t 系统中的顾客人数，则有

$$n(t) = 0, \text{如果 } 0 \leqslant t < 4,$$

$$n(t) = 1, \text{如果 } 4 \leqslant t < 6,$$

$$n(t) = 0, \text{如果 } t = 6$$

□

6.3 两个服务员的串联排队系统

现在考虑有两个服务员的串联排队系统. 假设顾客到达时间服从非齐次泊松过程, 且每一位顾客必须先找服务员 1, 只有服务员 1 完成对该顾客的服务后, 才能找服务员 2. 称这样的系统为*串联的*或*序贯的*排队系统. 当服务员 1 空闲时, 到达的顾客可以即时得到服务员 1 的服务; 否则, 就要排队等候服务员 1 的服务. 类似地, 当顾客得到服务员 1 的服务后, 如果服务员 2 空闲, 则该顾客可以立即得到服务员 2 的服务, 否则, 他就要排队等候服务员 2 的服务. 当完成服务员 2 的服务后, 该顾客才离开系统. 以 $G_i, i=1,2$ 表示服务员 i 的服务时间的分布 (见图 6-2).

97

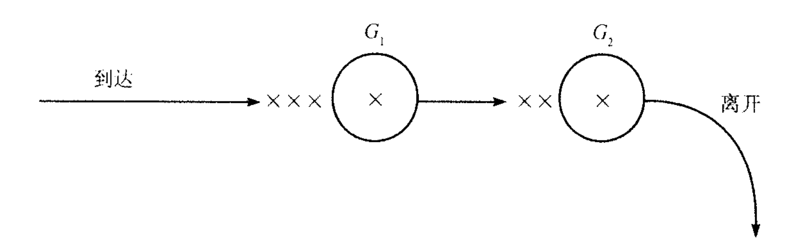


图 6-2 两个服务员的串联排队系统

假设我们感兴趣的是想利用模拟方法研究顾客得到服务员 1 和服务员 2 的总服务时间的分布. 我们将利用下列变量解决上述问题.

时间变量 t

系统状态变量SS

(n_1, n_2): 表示有 n_1, n_2 位顾客分别在等候或接受服务员 1 和服务员 2 的服务.

计数变量

N_A: 至时刻 t 到达的顾客数

N_D: 至时刻 t 离开的顾客数

输出变量

$A_1(n)$: 顾客 n 的到达时间 $(n \geqslant 1)$

$A_2(n)$: 顾客 n 到达服务员 2 处的时间 $(n \geqslant 1)$

$D(n)$: 顾客 n 的离开时间 $(n \geqslant 1)$

事件列表 t_A, t_1, t_2: t_A 表示下一顾客的到达时间, t_i 表示服务员 i 对正在接受服务的顾客的服务时间 $(i=1,2)$. 如果此时服务员 i 空闲, 则记 $t_i=\infty, i=1,2$. 事件列表始终包含三个变量 t_A, t_1, t_2.

为开始进行模拟, 我们取变量和事件列表的初始值如下:

初始化

取 $t = N_A = N_D = 0$.

取 $\mathbf{SS} = (0, 0)$.

98

生成 T_0，且取 $t_A = T_0, t_1 = t_2 = \infty$.

为更新此系统，我们就要改变时间直至下一个事件的出现. 至于出现哪个事件，则要考虑多种情形，且依赖于事件列表中的较小者. 下面以 Y_i 记分布为 G_i 的服务时间随机变量 $(i = 1, 2)$.

$$\mathbf{SS} = (n_1, n_2), \quad \mathbf{EL} = t_A, t_1, t_2$$

情形 1：$t_A = \min(t_A, t_1, t_2)$

重新取 $t = t_A$.

重新取 $N_A = N_A + 1$.

重新取 $n_1 = n_1 + 1$.

生成 T_t 并重新取 $t_A = T_t$.

如果 $n_1 = 1$，则生成 Y_1 且重新取 $t_1 = t + Y_1$.

收集输出数据 $A_1(N_A) = t$.

情形 2：$t_1 < t_A, t_1 \leqslant t_2$

重新取 $t = t_1$.

重新取 $n_1 = n_1 - 1, n_2 = n_2 + 1$.

如果 $n_1 = 0$，重新取 $t_1 = \infty$；否则，生成 Y_1 且重新取 $t_1 = t + Y_1$.

如果 $n_2 = 1$，则生成 Y_2 且重新取 $t_2 = t + Y_2$.

收集输出数据 $A_2(N_A - n_1) = t$.

情形 3：$t_2 < t_A, t_2 < t_1$

重新取 $t = t_2$.

重新取 $N_D = N_D + 1$.

重新取 $n_2 = n_2 - 1$.

如果 $n_2 = 0$，则重新取 $t_2 = \infty$.

如果 $n_2 > 0$，则生成 Y_2 且重新取 $t_2 = t + Y_2$.

收集输出数据 $D(N_D) = t$.

利用上述更新方法，我们很容易得到模拟此系统和收集相关数据的程序.

6.4　两个服务员的并联排队系统

现考虑模拟有两个服务员的排队系统. 如果两个服务员均忙碌，则到达的顾客排队等候. 如果服务员 1 空闲，则顾客接受服务员 1 的服务；如果服务员 2 空闲，

则接受服务员 2 的服务. 当顾客得到服务后（不论是服务员 1 还是服务员 2），则离开系统，且下一个等候时间最久的顾客接受服务. 设服务员 i 的服务时间分布为 $G_i, i = 1, 2$(见图 6-3).

99

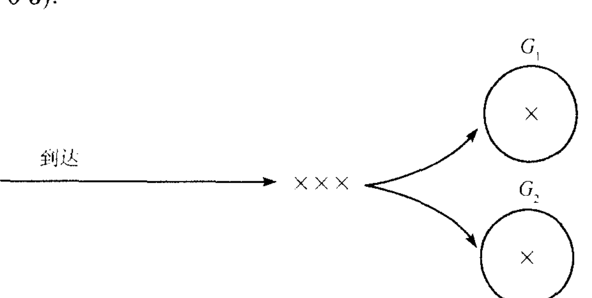


图 6-3 两个服务员的并联排队系统

假设我们想模拟上述模型，且想知道每位顾客在服务站的时间、每位服务员服务的顾客人数等. 由于有两位服务员为顾客服务，故顾客离开的顺序不一定和其到达顺序相同. 因此，为掌握哪位顾客结束服务离开系统，就必须知道哪些顾客在系统中. 于是，我们按照顾客到达时间给顾客编号：记先到的为顾客 1，第二位为顾客 2 等等. 由于顾客是以其到达时间的顺序接受服务的，故知道哪位顾客正在接受服务以及有多少位正在排队等候服务这些信息，就能帮助我们识别等候服务的顾客. 假设顾客 $i, j(i < j)$ 正在接受服务，且共有 $n - 2 > 0$ 个顾客在排队. 由于编号小于 j 的顾客先于 j 接受服务，且没有编号大于 j 的顾客接受服务 (否则他们将先于 i 或 j 接受服务)，故 $j + 1, \cdots, j + n - 2$ 位顾客在排队等候服务.

我们将利用下述变量分析此系统：

时间变量 t

系统状态变量 (SS)

(n, i_1, i_2)：n 表示此系统中的顾客数，i_1, i_2 分别表示接受服务员 1，服务员 2 服务的顾客数. 当系统为空时，$\mathbf{SS} = (0)$，若唯一的顾客 j 接受服务员 1 或服务员 2 服务时，相应地 $\mathbf{SS} = (1, j, 0)$ 或 $\mathbf{SS} = (1, 0, j)$.

计数变量

N_A：至时刻 t 到达的顾客数

C_j：至时刻 t 由服务员 j 服务的顾客数 $(j = 1, 2)$

输出变量

$A(n)$：顾客 n 的到达时间 $(n \geqslant 1)$

$D(n)$：顾客 n 的离开时间 $(n \geqslant 1)$

事件列表 t_A, t_1, t_2：t_A 表示下一位顾客的到达时间，t_i 表示服务员 i 对正在接受服

100

务顾客的服务时间 $(i=1,2)$. 如果服务员 i 空闲, 则取 $t_i=\infty(i=1,2)$. 事件列表始终包含三个变量 t_A, t_1, t_2.

为开始进行模拟, 我们取变量的事件列表的初始值如下:

初始化

取 $t=N_A=C_1=C_2=0$.

取 $\mathbf{SS}=(0)$.

生成 T_0 并取 $t_A=T_0, t_1=t_2=\infty$.

为更新此系统, 我们就要改变时间直至下一个事件出现. 至于出现哪个事件, 则要考虑多种情形, 且依赖于事件列表中的较小者. 下面以 Y_i 记分布为 G_i 的服务时间随机变量 $(i=1,2)$.

情形 1: $\mathbf{SS}=(n,i_1,i_2), t_A=\min(t_A,t_1,t_2)$

重新取 $t=t_A$.

重新取 $N_A=N_A+1$.

生成 T_t 并重新取 $t_A=T_t$.

收集输出数据 $A(N_A)=t$.

如果 $\mathbf{SS}=(0)$:

重新取 $\mathbf{SS}=(1,N_A,0)$.

生成 Y_1 并重新取 $t_1=t+Y_1$.

如果 $\mathbf{SS}=(1,j,0)$:

重新取 $\mathbf{SS}=(2,j,N_A)$.

生成 Y_2 并重新取 $t_2=t+Y_2$.

如果 $\mathbf{SS}=(1,0,j)$:

重新取 $\mathbf{SS}=(2,N_A,j)$.

生成 Y_1 并重新取 $t_1=t+Y_1$.

如果 $n>1$:

重新取 $\mathbf{SS}=(n+1,i_1,i_2)$.

情形 2: $\mathbf{SS}=(n,i_1,i_2), t_1<t_A, t_1\leqslant t_2$

重新取 $t=t_1$.

重新取 $C_1=C_1+1$.

收集输出数据 $D(i_1)=t$.

如果 $n=1$:

101　重新取 $\mathbf{SS}=(0)$.

重新取 $t_1=\infty$.

如果 $n=2$:

重新取 $\mathbf{SS} = (1, 0, i_2)$.

重新取 $t_1 = \infty$.

如果 $n > 2$: 记 $m = \max(i_1, i_2)$ 且

重新取 $\mathbf{SS} = (n-1, m+1, i_2)$.

生成 Y_1 并重新取 $t_1 = t + Y_1$.

情形 3: $\mathbf{SS} = (n, i_1, i_2), t_2 < t_A, t_2 < t_1$

情形 3 的更新留作习题.

如果我们利用上述方法模拟此系统，且在某一事先给定的时间点停止模拟，则由输出变量和 C_1, C_2 的最终值可以得到每位顾客的到达和离开时间以及每个服务员的服务人数.

6.5 仓储模型

现考虑某商店存储单价为 r 的某类产品. 假设需求此产品的顾客量服从参数为 λ 的泊松过程，且每位顾客的需求量是一分布为 G 的随机变量. 为满足需要，店主手中必须有一定存量的产品，但当手中的库存较少时，店主就要从批发商订购一定数量的产品. 店主应用的策略被称为 (s, S) 策略：当手中的存量小于 s 且当前没有待处理的订单，则店主需订购一定数量的产品以使存量达到 $S(s < S)$. 也就是说，如果当前没有待处理的订单且存量 $x < s$，则需订购总量为 $S - x$ 的产品. 设订购 y 单位产品的费用为一给定函数 $c(y)$，且需 L 个时间单元才能得到这批产品. 另外，假设单位产品每单位时间的存储费用为 h. 当需求量多于商店的当前存量时，则商店卖出所有的产品，但将失去多余的订单.

下面我们将利用模拟方法来估计到某时间 T 时此商店的期望收益. 为了模拟，我们先定义如下变量和事件：

时间变量 t

系统状态变量 (x, y)　x 表示手头的存量，y 表示已订购但未交货的数量. 102

计数变量

C：至时刻 t 的总订购费用

H：至时刻 t 的总仓储费用

R：至时刻 t 的总收入

事件包含顾客和订单的到达时间，其时间为

t_0：下一顾客的到达时间

t_1：订购产品被送达的时间. 如果没有待处理订单，则 $t_1 = \infty$.

变量更新的完成依赖于事件出现的时间. 如果以 t 记当前时间且我们有上述变量值, 则考虑如下两种情形:

情形 1: $t_0 < t_1$

重新取 $H = H + (t_0 - t)xh$[其原因是我们要支付时间 t 与 t_0 间的每单位产品 $(t_0 - t)h$ 的存储费用].

重新取 $t = t_0$.

生成 D(一个分布为 G 的随机变量, 且为时间 t_0 时顾客的需求量).

记 $w = \min(D, x)$ 为商店卖出的产品总量. 则剩余的库存量为 $x - w$.

重新取 $R = R + wr$.

重新取 $x = x - w$.

如果 $x < s$, 且 $y = 0$, 则重新取 $y = S - x, t_1 = t + L$.

生成 U 并重新取 $t_0 = t - \frac{1}{\lambda}\ln(U)$.

情形 2: $t_1 \leqslant t_0$

重新取 $H = H + (t_1 - t)xh$.

重新取 $t = t_1$.

重新取 $C = C + c(y)$.

重新取 $x = x + y$.

重新取 $y = 0, t_1 = \infty$.

我们容易利用上述更新方法编写分析此模型的模拟程序, 并运行此模拟程序直至在某一事先给定时间 T 后第一个事件出现. 我们用 $(R - C - H)/T$ 作为商店单位时间内平均收益的估计. 改变 s 和 S 的值以确定商店最佳的存储策略.

6.6　保险风险模型

假设某保险公司的投保客户独立地向公司索赔且服从参数为 λ 的泊松过程. 另 103 外, 客户的索赔额是一分布 F 的随机变量. 再假设来此保险公司投保的新客户数服从参数为 ν 的泊松过程, 而留下来的老客户服从参数为 μ 的指数分布. 最后, 再假设每个客户单位时间内付给保险公司的费用 c 是固定的. 以 n_0 记最初的客户数, $a_0 \geqslant 0$ 表示最初资金, 我们感兴趣的是想利用模拟方法估计到时间 T 此公司的资金总是非负的概率.

为模拟此模型, 我们定义变量和事件如下:

时间变量 t

系统状态变量 (n, a): n 为客户数, a 是公司当前的资金.

事件有三种类型：新增一个客户，失去一个客户和索赔. 事件列表仅有一个值，即下一个事件出现的时间.

EL t_E

由 2.9 节所讲的关于指数分布的结果，我们能够说明事件列表为什么仅包含下一事件出现的时间. 如果 (n, a) 是时间 t 的系统状态，则由于独立指数随机变量的最小者仍是指数的，故下一事件出现的时间为 $t+X$，其中 X 是一参数为 $\nu+n\mu+n\lambda$ 的指数随机变量. 另外，不论下一事件何时出现，它都将是如下情况之一：

以概率 $\dfrac{\nu}{\nu+n\mu+n\lambda}$ 新增一个客户；

以概率 $\dfrac{n\mu}{\nu+n\mu+n\lambda}$ 失去一个客户；

以概率 $\dfrac{n\lambda}{\nu+n\mu+n\lambda}$ 索赔.

当确定下一个事件何时出现后，我们生成一个随机数以确定此事件将导致哪种情况出现，然后利用此信息重新确定系统状态值.

对于给定的状态变量 (n, a)，下面以 X 表示参数为 $\nu+n\mu+n\lambda$ 的指数随机变量；以 J 表示以概率 $\frac{\nu}{\nu+n\mu+n\lambda}$ 取 1，以概率 $\frac{n\mu}{\nu+n\mu+n\lambda}$ 取 2，以概率 $\frac{n\lambda}{\nu+n\mu+n\lambda}$ 取 3 的随机变量；以 Y 表示分布函数为 F 的索赔随机变量.

输出变量 I 为

$$I = \begin{cases} 1 & \text{如果公司资金在 } [0, t] \text{ 内非负,} \\ 0 & \text{其他情况.} \end{cases}$$

104

初始化

先取

$$t = 0, \quad a = a_0, \quad n = n_0$$

然后生成 X 并取

$$t_E = X$$

为更新此系统，我们向前移动至下一事件，并检验它是否超过时间 T.

更新步骤

情形 1：$t_E > T$：

取 $I = 1$ 且结束此次模拟.

情形 2：$t_E \leqslant T$：

重新取

$a = a + nc(t_E - t)$

$t = t_E$

生成 J：

$J=1$：重新取 $n=n+1$

$J=2$：重新取 $n=n-1$

$J=3$：生成 Y. 如果 $Y>a$，则取 $I=0$ 且结束此次模拟；否则重新取 $a=a-Y$

生成 X：重新取 $t_E=t+X$

连续重复上述更新步骤直至此次模拟完成.

6.7 维修问题

一个系统的运转需要 n 台机器正常工作，为预防故障机器影响系统运转，现有几台可用机器备用. 只要某台机器出现故障，就立即用备用机器替换它，且把它送到修理工厂进行维修，而修理厂仅有一位修理人员且一次仅能维修一台故障机器. 105 一旦故障机器被修好，则立即转入备用状态（见图 6-4）. 设每次维修时间均是独立的分布函数为 G 的随机变量，而每台机器投入使用后，其正常工作时间是一独立于过去的随机变量，且其分布函数为 F.

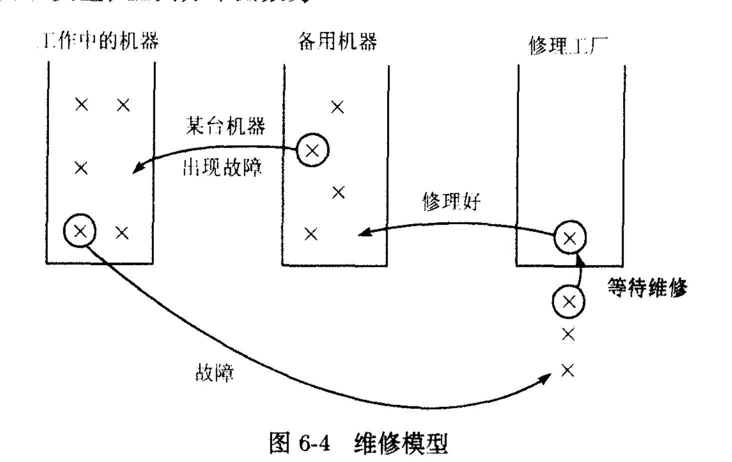


图 6-4 维修模型

当有一台机器出现故障而没有可用的备用机器时，我们称此系统“崩溃”. 假设我们最初利用 $n+s$ 台可用机器中的 n 台，而其余的 s 台备用，且感兴趣的是模拟此系统并近似 $E[T]$，其中 T 为系统崩溃时间.

为模拟上述模型，我们应用下述变量.

时间变量 t

系统状态变量 r：时间 t 发生故障的机器数

由于正在工作的机器出现故障或某台故障机器维修完成均改变系统状态变量，故

我们称这两种情况为“事件”. 为了知道下一事件何时发生，我们必须知道现正常工作的机器发生故障的时间和正在维修的机器所需的维修时间（如果只有一台机器在维修）. 因为我们总需要确定 n 个故障时间中的最小者，故用一个由小到大排列的列表来存储这 n 个时间最方便. 于是，我们取事件列表如下：

$$\text{事件列表：} t_1 \leqslant t_2 \leqslant t_3 \leqslant \cdots \leqslant t_n, t^*$$

其中 $t_1, \cdots, t_n$ 是 n 台工作机器发生故障的时间，t^* 是正在维修机器被修好的时间（如果没有机器在维修，则取 $t^* = \infty$）.

106

为开始模拟，我们取各变量的初始值如下：

初始化

取 $t = r = 0, t^* = \infty$.

生成来自分布函数 F 的 n 个独立随机数 $X_1, \cdots, X_n$.

对 $\{t_i\}$ 排序，使 $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_i \leqslant \cdots \leqslant t_n$.

取事件列表：$t_1, \cdots, t_n, t^*$.

根据如下两种情形更新系统：

情形 1：$t_1 < t^*$

重新取 $t = t_1$.

重新取 $r = r + 1$(其原因是另一台机器出现故障).

如果 $r = s + 1$，停止模拟并收集数据 $T = t$(其原因是此时有 $s + 1$ 个机器发生故障且没有备用的机器可用).

如果 $r < s + 1$，则生成来自分布函数 F 的随机数 X. 此随机数将表示备用机器投入使用后的工作时间. 现将值 $t_2, t_3, \cdots, t_n, t + X$ 由小到大重新排序，且以 t_i 记第 i 个最小的 $(i = 1, \cdots, n)$.

如果 $r = 1$，生成来自分布函数 G 的随机数 Y 且取 $t^* = t + Y$(其原因是此时发生故障的机器仅有一台，故它将得到及时的维修，而 Y 将是它的维修时间，且维修将在 $t + Y$ 时刻完成).

情形 2：$t^* \leqslant t_1$

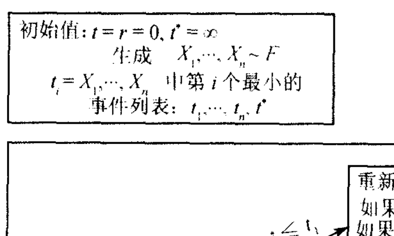
重新取 $t = t^*$.

重新取 $r = r - 1$.

如果 $r > 0$，则生成来自分布函数 G 的随机数 Y(它表示刚得到维修服务的机器的维修时间) 且重新取 $t^* = t + Y$.

如果 $r = 0$，取 $t^* = \infty$.

上述更新方法的图示说明见图 6-5.



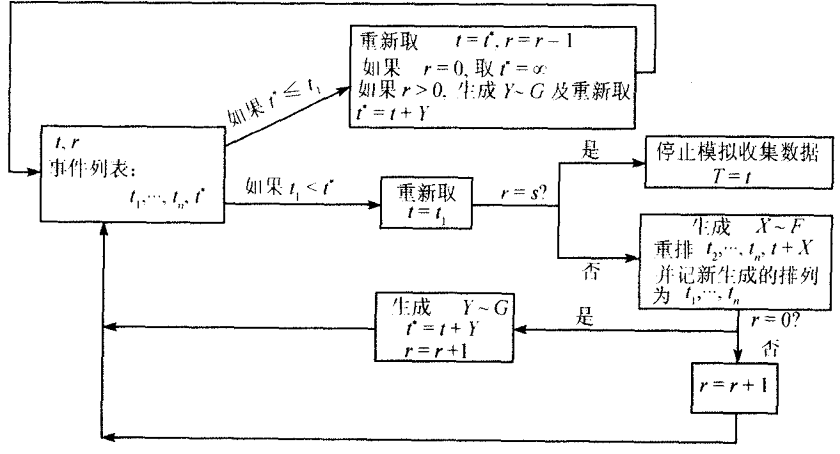


图 6-5 维修问题的图示说明

每次停止（当 $r = s + 1$ 时）就表明一次模拟的完成，且得到的输出值是崩溃时间 T. 之后，我们再进行初始值选取和下一次模拟. 这样，我们总共进行 k 次模拟，得到输出值为 $T_1, \cdots, T_k$. 因为这 k 个随机变量是独立的，且每一个均反映系统崩溃时间，故其平均值 $\sum_{i=1}^{k} T_i/k$ 即为平均崩溃时间 $E[T]$ 的估计. 第 7 章将考虑何时停止模拟，即 k 的选取问题. 并且，应用第 7 章中的方法对模拟结果进行统计分析.

107

6.8 股票期权的模拟

以 $S_n, n \geqslant 0$ 记第 n 天某指定股票的收盘价. 一个通用的模型为

$$S_n = S_0 \exp\{X_1 + \cdots + X_n\}, \quad n \geqslant 0$$

其中 $X_1, X_2, \cdots$ 是独立的来自均值为 μ, 方差为 σ^2 的正态随机量列. 此模型假设每个交易日相对于前一个交易日的增长率具有相同的分布，我们称之为**对数正态随机游动模型**. 记 $\alpha = \mu + \sigma^2/2$. 现假设你有权在以后 N 天内的任何一天以每股固定价 K(称之为**敲定价**) 购买此种股票. 如果当股票价格为 S 时你执行此期权，则由于你仅需支付 K 元，故你的获利将为 $S - K$(从理论上讲，由于你执行此期权后，转身就能以价格 S 卖出). 拥有此期权的平均获利依赖于你所应用的期权执行策略 (如果在这 N 天之内，股票价格没有超过 K，则显然你不会执行此期权). 现已证

明，当 $\alpha \geqslant 0$ 时，最佳策略是：在持有期间，如果股票价格超过 K，则执行此期权；否则，不执行此期权. 因为 $X_1+\cdots+X_N$ 是一均值为 $N\mu$、方差为 $N\sigma^2$ 的正态随机变量，故容易清楚地计算此策略的收益. 然而，当 $\alpha < 0$ 时，要刻画此种期权的最佳或近似最佳策略是非常困难的，并且我们也很难清楚地计算一个好的合理策略的平均获利. 下面，我们将给出一个当 $\alpha < 0$ 时可以利用的策略. 尽管此策略远不是最佳的，但看来是一个好的合理策略. 此策略为：当离期权到期日还有 m 天时，如果对于任一个 $i=1,\cdots,m$，执行期权得到的期望收益大于等 i 天后再执行期权的收益（即等到第 i 天若股票的价格高于 K，则执行期权；否则，则放弃此权力），则执行此期权.

108

以 $P_m = S_{N-m}$ 记期权到期前第 m 天的股票价. 我们建议的策略为：

策略 假设距期权到期日还剩 m 天，如果此时

$$P_m > K$$

且对于每个 $i=1,\cdots,m$,

$$P_m > K + P_m \mathrm{e}^{i\alpha}\Phi(\sigma\sqrt{i}+b_i) - K\Phi(b_i)$$

则执行此期权，其中

$$b_i = \frac{i\mu - \ln(K/P_m)}{\sigma\sqrt{i}}$$

$\Phi(x)$ 为标准正态分布函数，且可由下式较精确地近似计算：当 $x \geqslant 0$ 时

$$\Phi(x) \approx 1 - \frac{1}{\sqrt{2\pi}}(a_1y + a_2y^2 + a_3y^3)\mathrm{e}^{-x^2/2}$$

当 $x<0$ 时，$\Phi(x) = 1-\Phi(-x)$，其中

$$y = \frac{1}{1+0.33267x},$$
$$a_1 = 0.4361836,\quad a_2 = -0.1201676,\quad a_3 = 0.9372980$$

如果执行此期权，则以 SP 记执行此期权时股票的价格；如果此期权没有被执行，则取 SP 等于 K. 为确定上述策略的期望收益，即 $E[SP]-K$，我们必须借助于统计模拟. 对于给定的参数 μ,σ,N,K,S_0，我们先生成均值为 μ，标准差为 σ 的正态随机数，之后利用关系式

109

$$P_{m-1} = P_m\mathrm{e}^X$$

来产生 P_m. 如果 P_m 为期权到期前第 m 天的股票收盘价，且此时的策略是不行使此期权，则我们再产生 X 并得到 P_{m-1}，且利用计算机检验此时是否执行此期权.

如果决定执行，则在此次模拟中 $SP = P_{m-1}$；如果决定不执行，则继续确定下一个交易日的股票收盘价. 以此类推，多次随机模拟后，$SP - K$ 的平均值即为此种期权的上述策略之期望收益的估计值.

6.9　模拟模型的验证

由于离散事件模拟方法是用计算机程序来实现的，人们为确保此程序无缺陷，当然会用各种标准方法以消除种种缺陷. 下面，我们将讨论几种可专门用来消除模拟模型缺陷的方法.

对于一个大程序而言，人们将努力去除模块或子程序中的缺陷，其方法通常是把程序分成几个小的易于处理的部分（每部分从逻辑上都自成一体），之后去除每部分的缺陷. 例如，在统计模拟中，可以先把随机变量的生成作为一个模块，之后再分别去除每一模块的缺陷.

一般地，统计模拟程序总是要利用很多输入变量，而这些变量的适当选取可以把一个模拟模型简化成一个有解析解的或前面已广泛研究过的模型，于是，我们可以对模拟结果和真实结果进行比较.

在检验阶段，程序应输出它产生的所有随机变量值，并适当选取某些简单的特殊情况，比较模拟结果和手工结果是否有差异. 例如，假设我们想模拟有 k 个服务员前 T 个时间单元内的排队系统，而取 $T = 8$(或另一个较小值) 和 $k = 2$. 此时模拟程序生成如下数据：

顾客数:	1	2	3	4	5	6
到达时刻:	1.5	3.6	3.9	5.2	6.4	7.7
服务时刻:	3.4	2.2	5.1	2.4	3.3	6.2

110

并假设程序输出的六位顾客在本系统的平均时间为 5.12.

然而，通过手工计算可知：第一位顾客花在系统中的时间为 3.4；第二位顾客为 2.2（注意只有两位服务员）；第三位顾客到达时刻为 3.9，得到服务的时刻为 4.9（当第一位顾客离开后），而其需要的服务时间为 5.1，故他花在系统中的时间为 6.1；第四位顾客的到达时刻为 5.2，得到服务的时刻为 5.8（当第二位顾客离开），而经过时间 2.4 后离开，故他在系统中的时间为 3.0. 以此类推，手工计算的结果为：

到达时刻:	1.5	3.6	3.9	5.2	6.4	7.7
开始服务时刻:	1.5	3.6	4.9	5.8	8.2	10.0
离开时刻:	4.9	5.8	10.0	8.2	11.5	16.2
在系统中的时间:	3.4	2.2	6.1	3.0	5.1	8.5

于是，时刻 $T=8$ 之前来到此系统的顾客在系统中的平均时间为

$$\frac{3.4+2.2+6.1+3.0+5.1+8.5}{6}=4.71666\cdots$$

它与由程序得到的 5.12 有误差.

利用跟踪器查找计算机程序中的错误是非常有效的. 每当一个事件出现后，系统状态变量、事件列表和计数变量均被记录在跟踪器中. 由跟踪器内容可以判断模拟系统的表现是否如愿 (如果跟踪器显示没有明显误差，则人们可以检验有关输出变量的计算).

习 题

1. 编写程序以生成 6.2 节模型的期望输出. 应用此程序估计顾客在系统中的平均时间和服务员的平均加班时间，其中顾客到达过程是一个参数为 10 的泊松过程，服务时间的密度函数为

$$g(x)=20\mathrm{e}^{-40x}(40x)^2,\ x>0$$

$T=9$. 先模拟 100 次，之后再模拟 1000 次.

111

2. 对于 6.2 节的模型，假设我们还希望得到服务员一天中空闲时间的有关信息. 请解释如何完成此任务.
3. 假设到达一个单服务员排队系统的顾客服从一个非齐次泊松过程，开始时的参数为每小时 4 个；在后面的 5 小时内，其参数稳定增长直至每小时 19 个；而在后面的 5 小时内，其参数又稳定地降至每小时 4 个. 如此反复，即 $\lambda(t+10)=\lambda(t)$. 假设服务时间分布是一参数为每小时 25 的指数分布. 只要服务员完成一次服务且没有顾客等候，则假设他将休息一次，其休息时间服从 (0,0.3) 上的均匀分布. 当服务员休息之后，如仍没有顾客等候，则他将再休息一次. 请基于 500 次模拟估计服务员在前 100 个小时内的平均休息时间.
4. 写出 6.4 节模型之情形 3 的更新方案.
5. 考虑一个单服务员的排队模型，其中到达顾客服从一个非齐次泊松过程. 如果服务员空闲，则到达的顾客即时得到服务；否则，就要排队等候. 现假设每位顾客能排队等候的时间为分布 F 的随机变量，否则就离开系统. 以 G 表示服务时间分布. 如一位顾客在得到服务之前离开，则称损失一位顾客. 现假设我们希望估计时间 T 内损失的平均顾客数. 请定义相应变量和事件，并给出更新步骤.
6. 在习题 5 中，假设到达过程是一参数为 5 的泊松过程；F 为 (0,5) 上的均匀分布；G 是一参数为 4 的指数分布. 基于 500 次模拟估计时间 100 内损失的平均顾客数 (假设先到的顾客先得到服务).

7. 重复习题 6，但此时假设每当服务员完成一次服务后，不是为排队时间最长的顾客服务，而是为最着急的顾客服务. 也就是说，假设现有两位顾客排队等候，且他们最多能等候的时间分别为 t_1 和 t_2（即如二位的等候时间若超过 t_1 和 t_2 就将分别离开），如果 $t_1 < t_2$，则服务员先为第一位服务；否则就先为第二位服务. 你认为这种服务方式会增加还是减少损失的平均顾客数？

8. 对于 6.4 节中的模型，假设 G_1 是参数为 4 的指数分布，G_2 是参数为 3 的指数分布，到达顾客服从参数为 6 的泊松过程. 编写模拟程序，产生前 1000 位到达顾客的相应数据，并由此估计：

112

   (a) 这些顾客花在系统中的平均时间;
   (b) 由服务员 1 服务的顾客比例;
   (c) 再做一次模拟，并由此回答 (a) 和 (b)，且比较两次模拟的结果.

9. 假设在 6.4 节的两服务员并联排队系统中，共有两队在分别等候两个服务员的服务，且新来的顾客排在最短的那支队伍等候. 如果两支队伍一样长或两个服务员均空闲，则新来的顾客等候服务员 1 的服务.
   (a) 适当选取分析此模型的变量和事件，并给出更新步骤.
   如假设与习题 8 有相同的分布和参数，则求：
   (b) 前 1000 位顾客花在此系统中的平均时间.
   (c) 得到服务员 1 服务的顾客比例.
   在运行本程序前，你期望 (b) 和 (c) 的答案比习题 8 中的相应答案大还是小？

10. 在习题 9 中假设新来的顾客以概率 p 等候服务员 1 的服务且独立于其他变量.
   (a) 适当选取分析此模型的变量和事件，并给出更新步骤.
   (b) 应用习题 9 中的参数且取 p 为习题 9 中问题 (c) 的答案，模拟此系统并估计在习题 9 中之问题 (b) 中定义的量. 你期望本问题的答案比习题 9 得到的答案大还是小？

11. 假设向某保险公司的索赔客户服从参数为每天 10 个的泊松过程，索赔额是一个均值为 \$1000 的指数随机变量. 此保险公司以每天 \$11 000 的常数连续不断地收得报酬. 假设此公司的初始资金为 \$25 000 元，利用统计模拟估计此公司在前 365 天的资金始终为正的概率.

12. 对于 6.6 节中的模型，假定已知此公司的资金在时间 T 前变负. 现我们感兴趣的是它在何时资金变负及其资金缺口量. 请解释如何利用给定的统计模拟方法得到相关的数据.

13. 对于 6.7 节中的维修模型：
   (a) 编写模拟此模型的程序.

113

   (b) 利用此程序估计当 $n=4, s=3, F(x)=1-\mathrm{e}^{-x}, G(x)=1-\mathrm{e}^{-2x}$ 时的崩溃时间.

14. 假设在 6.7 节中的模型中，维修工厂有两名服务员，且他们维修机器的时间均为分布 G 的随机变量. 请画出此系统的流程图.
15. 在撞击试验系统中，撞击服从参数为每小时 1 个的泊松过程，且每次撞击均给系统带来一定的损害. 假设损害是独立的，是密度函数为

$$f(x) = x\mathrm{e}^{-x},\ x > 0$$

的随机变量 (也假设它与撞击发生的时间独立)，且损害以指数 α 消退，即一个初始损害为 x 的撞击，当其发生 s 时间后，其残余损害为 $x\mathrm{e}^{-\alpha s}$. 另外，假设损害是可以累加的，即，如果有两次撞击，其初始损害分别为 x_1, x_2，发生时间分别为 t_1 和 t_2，则时间 t 的损害为 $\sum_{i=1}^{2} x_i \mathrm{e}^{-\alpha(t-t_i)}$. 当总的损害超过某固定值 C 时，此系统瘫痪.

(a) 假设我们感兴趣的是利用统计模拟来估计系统瘫痪的平均时间. 定义此模型的事件和变量，并给出如何模拟此模型的流程图.

(b) 编写能生成 k 次模拟结果的程序.

(c) 通过比较程序输出和手工计算结果，验证此程序.

(d) 取 $\alpha = 0.5, C = 5, k = 1000$，运行此程序并利用其输出估计此系统瘫痪的平均时间.

16. 假设到达某通讯装置的信号服从参数为每小时 2 个的泊松分布，且此装置共有三个频道. 当三个频道均没被占用时，信号可进入任一频道；但当三个频道均被占用时，此信号将丢失. 一个信号联系一个频道的时间是一随机变量，且依赖于信号到达时的天气状况. 特别地，如果信号到达时的天气状况 "良好"，则此信号的处理时间为一分布函数为

$$F(x) = x,\ 0 < x < 1$$

的随机变量；如果信号到达时的天气 "不好"，则其处理时间为一分布函数为

$$F(x) = x^3,\ 0 < x < 1$$

的随机变量. 假设开始时天气良好，之后在良好和不好之间周期变化，其中良好和不好天气的长度分别为 2 小时和 1 小时 (这就是说，在时间 5，天气由良好转为不好). [114]

假设我们感兴趣的是时间 $T = 100$ 时丢失信号数的分布.

(a) 定义事件和变量以便我们能利用离散事件方法.

(b) 写出上述方法的流程图.

(c) 编写上述方法的程序.

(d) 通过比较程序输出和手工计算结果，验证上述程序.

(e) 运行上述程序并估计前 100 个小时内丢失信号的平均数.

17. 假设某股票的当前价是 100，且某人拥有一张期权，此期权允许他在后面 20 天的任何一天以 100 的价格购买此股票. 利用 6.8 节中参数 $\mu = -0.05, \sigma = 0.3$ 的模型及策略，通过模拟方法估计此期权的期望收益.

参考文献

[1] Banks, J., and J. Carson, *Discrete-Event System Simulation.* Prentice-Hall, New Jersey, 1984.

[2] Clymer, J., *Systems Analysis Using Simulation and Markov Models.* Prentice-Hall, New Jersey, 1990.

[3] Gottfried, B., *Elements of Stochastic Process Simulation.* Prentice-Hall, New Jersey, 1984.

[4] Law, A. M., and W. D. Kelton, *Simulation Modelling and Analysis.* 3rd ed. McGraw-Hill, New York, 1997.

[5] Mitrani, I., *Simulation Techniques for Discrete Event Systems.* Cambridge University Press, Cambridge, U. K., 1982.

[6] Peterson, R., and E. Silver, *Decision Systems for Inventory Management and Production Planning.* Wiley, New York, 1979.

[7] Pritsker, A., and C. Pedgen, *Introduction to Simulation and SLAM.* Halsted Press, New York, 1979.

[8] Shannon, R. E., *Systems Simulation: The Art and Science.* Prentice-Hall, New Jersey, 1975.

115

[9] Solomon, S. L., *Simulation of Waiting Line Systems.* Prentice-Hall, New Jersey, 1983.

第 7 章　模拟数据的统计分析

引　　言

人们通常用模拟来研究确定与某特定随机模型有关的量 θ. 若通过系统模拟输出的数据 X 是一个期望为 θ 的随机变量. 再一次独立的模拟，或者说第二次运行此模拟得到一个新的均值为 θ 的独立随机变量. 如此进行，直到得到 k 个独立同分布的均值为 θ 的随机变量 $X_1, X_2, \cdots, X_k$. 这 k 个值的平均 $\bar{X} = \sum_{i=1}^{k} X_i/k$ 就作为 θ 的估计或近似.

本章考虑确定何时停止模拟即确定 k 的适当的取值问题. 考虑 θ 的估计的性质将有助于我们确定何时停止模拟. 另外，我们也将给出如何得到 θ 的满足给定置信度的置信区间的方法.

在本章的最后一节，我们将利用统计中很重要的统计方法，即，自助法来给出如何估计比样本均值更复杂的某些估计量的性质.

7.1　样本均值和样本方差

假设 $X_1, X_2, \cdots, X_n$ 是独立同分布的随机变量，用 θ 和 σ^2 分别表示它们的均值和方差，即 $\theta = E[X_i]$, $\sigma^2 = \text{Var}(X_i)$. 这 n 个数值的算术平均值

$$\bar{X} \equiv \sum_{i=1}^{n} \frac{X_i}{n}$$

117

叫做样本均值. 当总体均值 θ 未知时，通常用样本均值估计它.

因为

$$E[\bar{X}] = E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \sum_{i=1}^{n} \frac{E[X_i]}{n} = \frac{n\theta}{n} = \theta \tag{7.1}$$

所以 $\bar{X}$ 是 θ 的一个无偏估计（如果一个参数的估计的期望值等于该参数，我们称该估计是这个参数的无偏估计）. 为了判断 $\bar{X}$ 作为总体均值 θ 的估计的“价值”，我们考虑它的均方误差，即 $\bar{X}$ 与 θ 的差的平方的期望值. 由于

$$
\begin{aligned}
E[(\bar{X}-\theta)^2] &= \mathrm{Var}(\bar{X}) && \text{因为} E[\bar{X}]=\theta \\
&= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) && (7.2)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) && \text{由独立性} \\
&= \frac{\sigma^2}{n}. && \text{因为}\mathrm{Var}(X_i)=\sigma^2
\end{aligned}
$$

因此，样本均值 $\bar{X}$ 是一个随机变量，它的期望是 θ，方差是 σ^2/n. 由于随机变量不太可能偏离它的均值太多个标准差（标准差等于方差的算术平方根），因此，当 $\sigma/\sqrt{n}$ 较小时，样本均值 $\bar{X}$ 是 θ 的一个好的估计.

注 就上面提到的随机变量不太可能偏离它的均值太多个标准差的根据是切比雪夫不等式和对模拟研究十分重要的中心极限定理. 对任意 $c>0$，切比雪夫不等式 (见第 2 章 2.7 节) 给出十分保守的界

118

$$
P\left\{|\bar{X}-\theta|>\frac{c\sigma}{\sqrt{n}}\right\}\leqslant\frac{1}{c^2}
$$

然而，根据中心极限定理，当 n 较大时，随机变量 $(\bar{X}-\theta)/(\sigma/\sqrt{n})$ 近似服从标准正态分布，而模拟研究的 n 值一般都很大，因此

$$
\begin{aligned}
P\left\{|\bar{X}-\theta|>\frac{c\sigma}{\sqrt{n}}\right\} &\approx P\{|Z|>c\}, \text{ 其中 } Z \text{ 是标准正态随机变量} \\
&= 2[1-\Phi(c)] \qquad (7.3)
\end{aligned}
$$

其中 Φ 是标准正态分布函数. 例如，由于 $\Phi(1.96)=0.975$，故等式 (7.3) 说明样本均值偏离 θ 的值超过 $1.96\sigma/\sqrt{n}$ 的概率近似为 0.05，然而较弱的切比雪夫不等式只能得到概率小于 $1/(1.96)^2=0.2603$. □

由于总体方差 σ^2 通常未知，因此直接利用 σ^2/n 的值作为标准来确定样本均值估计总体均值的好坏是有困难的. 因此，我们也需要对它进行估计. 由于

$$
\sigma^2=E[(X-\theta)^2]
$$

等于随机变量 X 与其期望之差的平方的期望，所以如同用 $\bar{X}$ 估计总体均值一样，我们自然用 $\sum_{i=1}^{n}(X_i-\bar{X})^2/n$ 估计总体方差 σ^2. 然而，为了使估计是无偏的 (还有其他技术原因) 我们更愿意用平方和 $\sum_{i=1}^{n}(X_i-\bar{X})^2$ 除以 $n-1$ 作为 σ^2 的估计.

定义 我们称由

$$
S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2
$$

定义的 S^2 为样本方差.

利用代数变形易得 (其证明留作习题)

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \tag{7.4}$$

我们容易证明样本方差是总体方差 σ^2 的无偏估计.

命题

$$E[S^2] = \sigma^2$$

119

证明 利用等式 (7.4), 我们得到

$$\begin{aligned}(n-1)E[S^2] &= E\left[\sum_{i=1}^{n} X_i^2\right] - nE[\bar{X}^2] \\ &= nE[X_1^2] - nE[\bar{X}^2]\end{aligned} \tag{7.5}$$

其中最后一个等式成立是因为所有的 X_i 分布相同. 我们注意到, 对任意随机变量 Y, 有 $\text{Var}(Y) = E[Y^2] - (E[Y])^2$, 或等价地有

$$E[Y^2] = \text{Var}(Y) + (E[Y])^2$$

于是, 我们得到

$$E[X_1^2] = \text{Var}(X_1) + (E[X_1])^2 = \sigma^2 + \theta^2$$

和

$$\begin{aligned}E[\bar{X}^2] &= \text{Var}(\bar{X}) + (E[\bar{X}])^2 \\ &= \frac{\sigma^2}{n} + \theta^2 \qquad \text{由 (7.2) 式和 (7.1) 式}\end{aligned}$$

因此, 根据等式 (7.5), 我们得到

$$(n-1)E[S^2] = n(\sigma^2 + \theta^2) - n\left(\frac{\sigma^2}{n} + \theta^2\right) = (n-1)\sigma^2$$

结论得证. □

我们用样本方差 S^2 作为总体方差 σ^2 的估计, 用样本标准差 $S(S = \sqrt{S^2})$ 作为标准差 σ 的估计.

在统计模拟中, 假设我们已连续得到模拟数据 X_i. 如果我们的目的是估计 $\theta = E[X_i]$ 的值, 那么我们应该何时停止产生新的数据呢? 这个问题的答案是: 首

先为估计的标准差选取一个可以接受的数值 d, 例如, 如果选取的 d 是估计 $\bar{X}$ 的标准差, 为了有 95% 的把握保证 $\bar{X}$ 偏离 θ 不超过 $1.96d$, 那么我们应该继续产生新的数据, 直到产生的数据个数 n 满足 $\sigma/\sqrt{n}$ 的估计 $S/\sqrt{n}$ 小于可以接受的数值 d. 当样本数较少时, 由于样本标准差 S 可能不是 σ 的一个特别好的估计 (正态近似也可能无效), 因此我们建议按下列程序确定何时停止产生新的数据.

120

确定何时停止产生新数据的方法

1. 为估计的标准差选择一个可以接受的数值 d.

2. 产生至少 100 个数据.

3. 持续产生新的数据, 直至数据个数 k 满足 $S/\sqrt{k} < d$, 其中 S 是基于这 k 个数据的样本标准差.

4. 用 $\bar{X} = \sum_{i=1}^{k} X_i/k$ 作为 θ 的估计.

例 7a　考虑一个下午五点以后不允许新到顾客进入的服务系统. 假定每天到达的顾客数服从同一概率分布, 我们感兴趣的是估计最后一个顾客离开系统的平均时间. 更进一步说, 假设我们想以至少 95% 的把握确信我们估计的结果偏离真值不超过 15 秒.

为满足上面的要求, 我们需要连续产生与最后一个顾客离开时间有关的数据 (每次均需通过模拟得到), 直到总共产生 k 个值, 其中 k 至少为 100 且满足 $1.96S/\sqrt{k} < 15$, 这里 S 是这 k 个数据的样本标准差 (以秒为单位). 最后一个顾客离开此系统的平均时间的估计即为这 k 个数据的平均值. □

如果我们能找到一种计算样本均值和样本方差的递推方法, 而不是当有了新数据后再重新用相同的方法计算样本均值和方差, 这将有助于我们应用上述方法, 确定何时停止产生新的数据. 下面我们将给出这种递推方法. 考虑一列数据 $X_1, X_2, \cdots$, 令

$$\bar{X}_j = \sum_{i=1}^{j} \frac{X_i}{j}$$

和

$$S_j^2 = \sum_{i=1}^{j} \frac{(X_i - \bar{X}_j)^2}{j-1}, \ j \geqslant 2$$

分别表示前 j 个数据的样本均值和样本方差. 下面的递推公式可用来连续地计算样本均值和样本方差.

记 $S_1^2 = 0, \bar{X}_0 = 0$,

121

$$\bar{X}_{j+1} = \bar{X}_j + \frac{X_{j+1} - \bar{X}_j}{j+1} \tag{7.6}$$

$$S_{j+1}^2 = \left(1-\frac{1}{j}\right)S_j^2 + (j+1)(\bar{X}_{j+1}-\bar{X}_j)^2 \tag{7.7}$$

□

例 7b 如果前三个数据分别是 $X_1=5, X_2=14, X_3=9$, 则根据等式 (7.6) 和 (7.7) 可得

$$\begin{aligned}\bar{X}_1 &= 5\\ \bar{X}_2 &= 5+\frac{9}{2}=\frac{19}{2}\\ S_2^2 &= 2\left(\frac{19}{2}-5\right)^2=\frac{81}{2}\\ \bar{X}_3 &= \frac{19}{2}+\frac{1}{3}\left(9-\frac{19}{2}\right)=\frac{28}{3}\\ S_3^2 &= \frac{81}{4}+3\left(\frac{28}{3}-\frac{19}{2}\right)^2=\frac{61}{3}\end{aligned}$$

□

当数据来自伯努利随机变量时，我们要估计的是概率，则上面的分析要做些修改. 假设我们可以产生随机变量 X_i，使得

$$X_i=\begin{cases}1 & \text{以概率}p\\ 0 & \text{以概率}1-p\end{cases}$$

且我们对估计 $E[X_i]=p$ 感兴趣. 此时，因为

$$\text{Var}(X_i)=p(1-p)$$

所以我们不需要用样本方差估计 $\text{Var}(X_i)$. 实际上，如果已经产生 n 个数据 $X_1,\cdots,X_n$，由于

$$\bar{X}_n=\sum_{i=1}^{n}\frac{X_i}{n}$$

是 p 的估计，则 $\bar{X}_n(1-\bar{X}_n)$ 自然就是 $\text{Var}(X_i)$ 的一个估计. 因此，在这种情况下，我们可以用如下的方法确定何时停止模拟.

1. 为估计的标准差选择一个可以接受的数值 d.
2. 产生至少 100 个数据.
3. 连续产生新数据，直到产生的数据个数 k 满足 $[\bar{X}_k(1-\bar{X}_k)/k]^{1/2}<d$ 为止.
4. p 的估计是这 k 个数据的平均值 $\bar{X}_k$.

例 7c　假设在例 7a 中，我们感兴趣的是估计到 5:30 仍有顾客在商店内的概率. 为此，记

$$X_i = \begin{cases} 1 & \text{如果在第}i\text{天 5:30 仍有顾客在商店内} \\ 0 & \text{其他情况} \end{cases}$$

我们将模拟至少 100 天，并且直到第 k 天，k 满足 $(p_k(1-p_k)/k)^{1/2} < d$ 为止，其中 $p_k = \bar{X}_k$ 是这 k 天中到 5:30 仍有顾客在商店内的比例，d 是为估计 p_k 的标准差选择的一个可以接受的数值. □

7.2　总体均值的区间估计

设 $X_1, X_2, \cdots, X_n$ 是独立同分布的随机变量，均值是 θ，方差是 σ^2. 尽管样本均值 $\bar{X} = \sum_{i=1}^n X_i/n$ 是 θ 的一个有效估计，但我们并不能期望 $\bar{X}$ 就真等于 θ，虽然它可以接近 θ. 若可以找到一个具体区间，且 θ 以一定的置信度包含在该区间内，则此区间有时将会更有意义.

为找到这样一个区间，需要知道估计量 $\bar{X}$ 的（近似）分布. 为得到 $\bar{X}$ 的（近似）分布，回顾等式 (7.1) 和 (7.2) 有

$$E[\bar{X}] = \theta \qquad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

因此，根据中心极限定理，当 n 充分大时，有

$$\sqrt{n}\frac{(\bar{X}-\theta)}{\sigma} \dot{\sim} N(0,1)$$

这里 $\dot{\sim} N(0,1)$ 表示"近似服从标准正态分布". 如果用样本标准差 S 代替未知的标准差 σ(根据著名的 Slutsky 定理)，则上述结论依然成立，也就是说，当 n 充分大时，有

$$\sqrt{n}\frac{(\bar{X}-\theta)}{S} \dot{\sim} N(0,1) \tag{7.8}$$

对任意的 α，$0<\alpha<1$，以 z_α 记满足下式的值

123

$$P\{Z > z_\alpha\} = \alpha$$

其中 Z 是服从标准正态分布的随机变量（例如 $z_{0.025} = 1.96$）. 由于标准正态分布的密度函数关于原点对称，则可知 Z 落在点 $z_{1-\alpha}$ 右侧部分的概率等于 $1-\alpha$，即 (见图 7-1) 有

$$z_{1-\alpha} = -z_\alpha$$

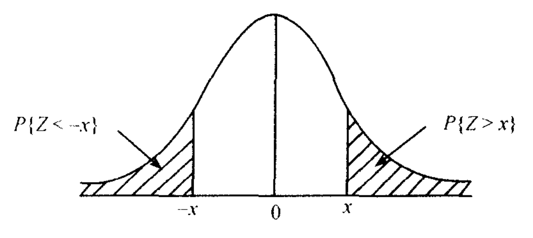


图 7-1 标准正态分布的密度

并且由图 7-1 亦可知

$$P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha$$

由 (7.8) 式得到

$$P\left\{-z_{\alpha/2} < \sqrt{n}\frac{(\bar{X} - \theta)}{S} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

乘以 −1 后，得到

$$P\left\{-z_{\alpha/2} < \sqrt{n}\frac{(\theta - \bar{X})}{S} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

它等价于

$$P\left\{\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \theta < \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right\} \approx 1 - \alpha \tag{7.9}$$

换句话说，总体均值 θ 将以概率 $1-\alpha$ 落在区间 $\bar{X} \pm z_{\alpha/2}S/\sqrt{n}$ 之中.

定义 如果样本均值和样本标准差的观测值分别是 $\bar{X} = \bar{x}$ 和 $S = s$，则区间 $\bar{x} \pm z_{\alpha/2}s/\sqrt{n}$ 叫做 θ 的置信度 (近似) 为 $100(1-\alpha)\%$ 的置信区间估计.

124

注 1 为了阐明“置信度为 $100(1-\alpha)\%$ 的置信区间”的含义，考虑 $\alpha = 0.05$，此时对应的 $z_{\alpha/2} = 1.96$. 在观测数据之前，样本均值是 $\bar{X}$，样本标准差是 S，θ 落在 $\bar{X} \pm 1.96S/\sqrt{n}$ 之间的概率 (近似) 等于 0.95 是对的. 在得到样本均值 $\bar{X}$ 观测值 $\bar{x}$ 和样本标准差 S 的观测值 s 之后，θ 是否在区间 $\bar{x} \pm 1.96s/\sqrt{n}$ 之中已不再涉及任何概率. 然而，在这次模拟中，我们以 95% 的把握相信 θ 是在这个区间 (因为我们知道经过多次模拟，得到大量的这类区间，确实有 95% 的这类区间包含均值 θ).

注 2 (一个技术上的附注.) 以上分析基于等式 (7.8)，即当 n 较大时，$\sqrt{n}(\bar{X} - \theta)/S$ 近似服从标准正态分布. 若原始数据 X_i 本身服从正态分布，则 $\sqrt{n}(\bar{X} - \theta)/S$ (确切) 服从自由度为 $n-1$ 的 t 分布. 据此，许多作者提出在一般情况下应用这个近似分布而并不要求原始分布一定是服从正态分布的. 然而，由于并不清楚在一般情况下自由度为 $n-1$ 的 t 分布的近似结果是否比正态分布的近似结果更好，又因

当 n 较大时，这两个分布近似相等，我们一般多用正态近似而不再引入一个 t 分布随机变量. □

在模拟研究中，现在考虑可产生新增加数据的情况. 问题是应该何时停止产生新的数据. 一种方法是：预先选取 α 和 l，然后产生新的数据，直至产生的数据使得 θ 的置信度为 $100(1-\alpha)\%$ 区间估计的长度小于 l. 由于该区间的长度是 $2z_{\alpha/2}S/\sqrt{n}$，则我们可以用下面的方法确定何时停止产生新的数据.

1. 产生至少 100 个数据.

2. 继续产生新的数据直至总共产生的数据个数 k，满足 $2z_{\alpha/2}S/\sqrt{k} < l$，其中 S 是基于这 k 个数据的样本标准差. [当新数据产生后，通过 (7.6) 和 (7.7) 的递推公式，数据 S 的值应不断更新.]

3. 如果 $\bar{x}$ 和 s 分别是 $\bar{X}$ 和 S 的观测值，则 θ 的置信度为 $100(1-\alpha)\%$ 的置信区间估计是 $\bar{x} \pm 2z_{\alpha/2}s/\sqrt{k}$，其长度小于 l.

技术注释 一般所用置信区间是在样本量确定的假设下得到的，而上面情况 125 的样本量明显是一个随机变量，其值依赖于产生的数据，统计水平较高的读者可能会对这点有所顾虑. 然而，当样本量较大时，我们可以证明上述方法是正确的，因此从模拟的角度看，我们可以放心地忽略这细微之处. □

沿用上节的记号，当 $X_1, \cdots, X_n$ 服从伯努利分布时，有

$$X_i = \begin{cases} 1 & \text{以概率}p \\ 0 & \text{以概率}1-p \end{cases}$$

在这种情况下，可以用 $\bar{X}(1-\bar{X})$ 估计 $\mathrm{Var}(X_i)$，将其应用于等式 (7.8)，当 n 较大时，

$$\sqrt{n}\frac{(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} \dot{\sim} N(0,1) \tag{7.10}$$

因此，对任意 α，

$$P\left\{-z_{\alpha/2} < \sqrt{n}\frac{(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} < z_{\alpha/2}\right\} = 1-\alpha$$

它等价于

$$P\left\{\bar{X} - z_{\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n} < p < \bar{X} + z_{\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}\right\} = 1-\alpha$$

因此，如果 $\bar{X}$ 的观测值是 p_n，则我们说 p 的"$100(1-\alpha)\%$ 的置信区间估计"是

$$p_n \pm z_{\alpha/2}\sqrt{p_n(1-p_n)/n}$$

7.3 估计均方误差的自助法

假设 $X_1, \cdots, X_n$ 是相互独立的随机变量，有共同的分布函数 F. 并假设我们感兴趣的问题是估计分布 F 的某些参数 $\theta(F)$. 例如，$\theta(F)$ 可能是本章上一节所述的分布 F 的均值、中位数或者方差，或者 F 的某些其他参数. 假设 $\theta(F)$ 的估计已经给出，记作 $g(X_1, \cdots, X_n)$. 为了判定是否值得用它作为 $\theta(F)$ 的估计，我们感兴趣的问题是估计它的均方误差. 也就是说，我们感兴趣的是估计下式的值：

126

$$\text{MSE}(F) \equiv E_F[(g(X_1, \cdots, X_n) - \theta(F))^2]$$

[其中使用记号 $\text{MSE}(F)$ 是为了不过分地把注意力放在估计 g 上，用记号 E_F 是为了表明是在随机变量都服从分布 F 的假设下求的期望.] 当 $\theta(F) = E[X_i]$ 和 $g(X_1, \cdots, X_n) = \bar{X}$ 时，S^2/n 是上面提到的 $\text{MSE}(F)$ 的一个估计，但我们并不是轻而易举就知道这一点. 下面我们给出一个有用的方法估计这个均方误差，该方法被称做自助法.

首先，我们注意到，如果分布函数 F 已知，则我们可以从理论上计算 θ 与其估计之间的偏差平方的期望；也就是说，我们可以计算均方误差. 然而，当我们观测到 n 个数据点的值以后，我们会更容易判断此分布看起来像什么分布. 实际上，假设数据的观测值是 $X_i = x_i, i = 1, \cdots, n$. 我们通常用被称做经验分布函数的 F_e 来估计此分布函数 F，即函数 $F_e(x)$ 是 $F(x)$ 的估计，其数值小于或等于 x 的概率正好等于 n 个数值中小于或等于 x 的数所占的比例，即

$$F_e(x) = \frac{\#\{i : X_i \leqslant x\}}{n}$$

考虑这个问题的另一种思路是认为 F_e 是随机变量 X_e 的分布函数，其中随机变量 X_e 等可能地取 $x_1, \cdots, x_n$ 这 n 个值中的任意一个. (如果 x_i 的值不是完全可区分的，则认为 X_e 等于 x_i 的概率等于 $x_1, \cdots, x_n$ 中等于 x_i 的数值的个数除以 n；也就是说，如果 $n = 3$, $x_1 = x_2 = 1, x_3 = 2$，则认为 X_e 是一个随机变量，等于 1,2 的概率分别是 2/3 和 1/3.)

如果 F_e "接近" F [事实上，强大数定律表明：当 $n \to \infty$ 时，$F_e(x)$ 以概率 1 收敛到 $F(x)$. 另外，Glivenko-Cantelli 定理指出这种收敛是以概率 1 对 x 一致收敛，故 n 较大时，F_e 应当接近 F]，则在某种意义下 $\theta(F_e)$ 将有可能接近 $\theta(F)$ (如 θ 是分布 F 的连续函数)，并且 $\text{MSE}(F)$ 应该近似等于

$$\text{MSE}(F_e) = E_{F_e}[(g(X_1, \cdots, X_n) - \theta(F_e))^2]$$

在上面的表述中，X_i 被当作独立随机变量，分布函数是 F_e. 数值 $\mathrm{MSE}(F_e)$ 称作 [127] $\mathrm{MSE}(F)$ 的*均方误差的自助法近似*.

为了获得对均方误差自助法近似的可行性分析，我们考虑一个用样本均值 $\bar{X}$ 估计总体均值的实例. (在这个例子中，没必要使用这个方法，因为已经有估计均方误差 $E(\bar{X}-\theta)^2=\sigma^2/n$ 的有效方法，即利用 S^2/n 的观测值估计 σ^2/n.)

例 7d　假设我们对用样本均值 $\bar{X}=\sum_{i=1}^{n}X_i/n$ 估计总体均值 $\theta(F)=E[X]$ 感兴趣. 如果观测值是 $x_i, i=1,\cdots,n$, 则经验分布函数 F_e 是对 $x_1,\cdots,x_n$ 的每一个点做 $1/n$ 的加权 (如果 x_i 不是完全可区分的，则联合加权). 因此 F_e 的均值是 $\theta(F_e)=\bar{x}=\sum_{i=1}^{n}x_i/n$, 均方误差的自助法估计 $\mathrm{MSE}(F_e)$ 由下式给出

$$\mathrm{MSE}(F_e)=E_{F_e}\left[\left(\sum_{i=1}^{n}\frac{X_i}{n}-\bar{x}\right)^2\right]$$

其中 $X_1,\cdots,X_n$ 是独立随机变量，各自分布为 F_e. 因为

$$E_{F_e}\left[\sum_{i=1}^{n}\frac{X_i}{n}\right]=E_{F_e}[X]=\bar{x}$$

可得

$$\mathrm{MSE}(F_e)=\mathrm{Var}_{F_e}\left(\sum_{i=1}^{n}\frac{X_i}{n}\right)=\frac{\mathrm{Var}_{F_e}(X)}{n}$$

因为

$$\begin{aligned}\mathrm{Var}_{F_e}(X)&=E_{F_e}[(X-E_{F_e}[X])^2]\\&=E_{F_e}\left[(X-\bar{x})^2\right]\\&=\frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right],\end{aligned}$$

则

[128]

$$\mathrm{MSE}(F_e)=\frac{1}{n^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

与常用的均方误差的估计 S^2/n 相比是相当地好. 实际上，因为 S^2/n 的观测值是 $\sum_{i=1}^{n}(x_i-\bar{x})^2/[n(n-1)]$, 故自助法近似与之相同. □

如果观测值是 $X_i=x_i, i=1,\cdots,n$, 则当经验分布函数 F_e 对 $x_1,\cdots,x_n$ 的每一个点做 $1/n$ 的加权时，一般容易计算出 $\theta(F_e)$ 的值. 例如，如果感兴趣的参数

$\theta(F)$ 是分布 F 的方差，则 $\theta(F_e) = \text{Var}_{F_e}(X) = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$. 要确定均方误差的自助法近似，我们需要计算

$$\text{MSE}(F_e) = E_{F_e}[(g(X_1, \cdots, X_n) - \theta(F_e))^2]$$

然而，上面的期望是在假设 $X_1, \cdots, X_n$ 相互独立，各自分布为 F_e 的条件下计算的，由于向量 $(X_1, \cdots, X_n)$ 可能取值是 $(x_{i_1}, x_{i_2} \cdots, x_{i_n}), i_j \in \{1, 2, \cdots, n\}, j = 1, \cdots, n$ 且取这 n^n 个值中每个的可能性是相等的，则

$$\text{MSE}(F_e) = \sum_{i_n} \cdots \sum_{i_1} \frac{[g(X_{i_1}, \cdots, X_{i_n}) - \theta(F_e)]^2}{n^n}$$

其中每个 i_j 的取值是从 1 到 n. 因此，在一般情况下，$\text{MSE}(F_e)$ 的计算需求 n^n 个项的和，即当 n 较大时，这是个无法完成的任务.

然而，据我们所知，有一个近似计算项数很多的平均值的有效方法，即统计模拟. 实际上，我们可以产生一个包含 n 个相互独立且分布函数为 F_e 的随机变量的集合 $\{X_1^1, \cdots, X_n^1\}$，并令

$$Y_1 = [g(X_1^1, \cdots, X_n^1) - \theta(F_e)]^2$$

接着，产生第二个集合 $\{X_1^2, \cdots, X_n^2\}$，并计算

$$Y_2 = [g(X_1^2, \cdots, X_n^2) - \theta(F_e)]^2$$

重复进行上述过程，直到获得 r 个变量 $Y_1, Y_2, \cdots, Y_r$. 因为 $Y_1, Y_2, \cdots, Y_r$ 是相互独立的随机变量且均值是 $\text{MSE}(F_e)$，据此，我们可以用它们的平均值 $\sum_{i=1}^{r} Y_i/r$ 作为 $\text{MSE}(F_e)$ 的估计.

注 1 很容易产生分布是 F_e 的随机变量 X，其原因是这样的随机变量等可能地取 $x_1, x_2 \cdots, x_n$ 中的每一个值. 于是，产生随机数 U，令 $X = x_I$，其中 $I = \text{Int}(nU) + 1$. (容易验证即使 x_i 不是完全可区分的，此法仍可行.)

129

注 2 上述的模拟方法允许我们对 $\text{MSE}(F_e)$ 做近似，因为它本身就是对所求的 $\text{MSE}(F)$ 的近似. 另外，已有人指出，大约 100 次模拟即取 $r = 100$ 就足够了. □

下例是自助法在分析排队模拟的输出结果上的应用.

例 7e 假设在例 7a 中，我们感兴趣的是估计长期以来一个顾客花费在系统中的平均时间. 以 W_i 记第 i 个顾客花费在系统中的时间 $(i \geqslant 1)$，我们感兴趣的量是

$$\theta \equiv \lim_{m \to \infty} \frac{W_1 + W_2 + \cdots + W_m}{m}$$

为证明上面的极限确实存在 (注意到随机变量 W_i 之间既不相互独立也不服从同一分布), 令 N_i 表示第 i 天到达的顾客数, 并令

$$D_1 = W_1 + \cdots + W_{N_1}$$

$$D_2 = W_{N_1+1} + \cdots + W_{N_1+N_2}$$

一般情况下, 当 $i > 2$ 时,

$$D_i = W_{N_1+\cdots+N_{i-1}+1} + \cdots + W_{N_1+\cdots+N_i}$$

换句话说, D_i 是第 i 天到达的所有顾客各自花费在系统中的时间总和. 我们可以将 θ 表示为

$$\theta = \lim_{m\to\infty} \frac{D_1 + D_2 + \cdots + D_m}{N_1 + N_2 \cdots + N_m}$$

因为此比值正是前 m 天到达的所有顾客各自花费在系统中的时间的平均值. 分子和分母分别除以 m, 得到

$$\theta = \lim_{m\to\infty} \frac{(D_1 + \cdots + D_m)/m}{(N_1 + \cdots + N_m)/m}$$

由于每天都服从相同的分布, 所以随机变量 $D_1, \cdots, D_m$ 独立同分布, 随机变量 $N_1, \cdots, N_m$ 也是独立同分布的. 因此, 根据强大数律, 前 m 个 D_i 的平均值以概率 1 收敛到它们共同的期望值, 对 N_i 也有类似的结论成立. 因此, 我们得到

130

$$\theta = \frac{E[D]}{E[N]}$$

其中 $E[N]$ 是一天中到达的顾客数的期望, $E[D]$ 是一天中到达的所有顾客各自花费在系统中的总时间的期望.

为估计 θ 的值, 我们可以模拟这个系统 k 天, 收集第 i 次模拟运行的数据 N_i, D_i, 其中 N_i 表示第 i 天到达的顾客数, D_i 表示第 i 天到达的顾客各自花费在系统中的时间总和 ($i = 1, \cdots, k$). 因为

$$\overline{D} = \frac{D_1 + D_2 + \cdots + D_k}{k}$$

可用来估计数量 $E[D]$,

$$\overline{N} = \frac{N_1 + N_2 + \cdots + N_k}{k}$$

可用来估计数量 $E[N]$, 因此可用

$$\theta = \frac{\overline{D}}{\overline{N}} = \frac{D_1 + \cdots + D_k}{N_1 + \cdots + N_k}$$

来估计数量 $\theta = E[D]/E[N]$. 应注意到该估计正好是前 k 天所有到达的顾客花费在系统中的平均时间.

我们使用自助法估计

$$\text{MSE} = E\left[\left(\frac{\sum_{i=1}^{k} D_i}{\sum_{i=1}^{k} N_i} - \theta\right)^2\right]$$

假设 D_i, N_i 的观测值分别是 $d_i, n_i, i = 1, \cdots, k$. 也就是说，我们的模拟结果是：在第 i 天到达 n_i 个顾客，他们在系统中总共花费时间是 d_i. 因此随机向量 (D, N) 的经验联合分布函数分配相等的权在这 k 个数对 $d_i, n_i, i = 1, \cdots, k$ 上. 即，由经验分布函数有

$$P_{F_e}\{D = d_i, N = n_i\} = \frac{1}{k}, \quad i = 1, \cdots, k$$

因此，

$$E_{F_e}[D] = \bar{d} = \sum_{i=1}^{k} \frac{d_i}{k}, \quad E_{F_e}[N] = \bar{n} = \sum_{i=1}^{k} \frac{n_i}{k}$$

131

并且有

$$\theta(F_e) = \frac{\bar{d}}{\bar{n}}$$

因此，

$$\text{MSE}(F_e) = E_{F_e}\left[\left(\frac{\sum_{i=1}^{k} D_i}{\sum_{i=1}^{k} N_i} - \frac{\bar{d}}{\bar{n}}\right)^2\right]$$

上面的计算假设 k 对随机向量 D_i, N_i 依分布 F_e 独立.

由于 $\text{MSE}(F_e)$ 的准确计算需要计算 k^k 项的和，故我们通过模拟实验求其近似值. 根据经验分布函数 F_e 产生 k 个独立的随机向量对 $D_i^1, N_i^1, i = 1, \cdots, k$, 并计算

$$Y_1 = \left(\frac{\sum_{i=1}^{k} D_i^1}{\sum_{i=1}^{k} N_i^1} - \frac{\bar{d}}{\bar{n}}\right)^2$$

然后，产生第二个集合 $D_i^2, N_i^2, i = 1, \cdots, k$, 并计算对应的 Y_2. 重复上述步骤，直到获得 r 个值 $Y_1, Y_2, \cdots, Y_r$ ($r = 100$ 应该足够了). 用这 r 个值的平均值 $\sum_{i=1}^{r} Y_i/r$

来估计 $\mathrm{MSE}(F_e)$ （一个顾客花费在系统中的平均时间的估计的均方误差），而 $\mathrm{MSE}(F_e)$ 本身是 $\mathrm{MSE}(F)$ 的一个估计. □

注　再生法　上面的分析是基于假设每天独立且服从相同概率分布得到的. 在实际应用中，同一概率分布描述的不是时间长度固定的天数，而是时间长度为随机的一个周期. 例如，考虑一个顾客到达是泊松过程的排队系统，并假设第一个顾客在 0 时刻到达. 如果随机时间 T 表示到达的顾客发现系统是空闲的下一个时刻，我们说从 0 到 T 的时间构成第一个周期. 第二个周期从 T 时刻开始直到到达的顾客发现系统是空闲的 T 时刻之后的第一个时间点，依此类推. 在多数模型中，容易发现遍及每个周期的过程的运动是独立同分布的. 因此，如果把一个周期当作一“天”，前面的所有分析仍然正确. 例如一个顾客花费在系统中的时间 θ，由 $\theta = E[D]/E[N]$ 给出，其中 D 表示一个周期内到达的顾客各自花费在系统中的时间总和，N 表示一个周期内到达的顾客人数. 如果我们产生 k 个周期的数据，θ 的估计仍是 $\sum_{i=1}^{k} D_i / \sum_{i=1}^{k} N_i$. 另外，如同上面一样，应用自助法同样可得到这个估计的均方误差的近似值.

132

通过模拟“周期”对系统进行分析的方法，叫做再生法. 所谓“周期”是指在该随机区间内过程遵从同一概率分布.

习　题

1. 对任意给定的数集 $\{x_1, \cdots, x_n\}$，利用代数运算证明

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

其中 $\bar{x} = \sum_{i=1}^{n} x_i/n$.

2. 设 X 是一个随机变量，取 $x_1, \cdots, x_n$ 中任意一个值的可能性均相等，应用等式 $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ 给出习题 1 的结论的概率证明.

3. 应用等式 (7.6) 和 (7.7) 给出的递推公式编写一个程序计算数据集的样本均值和样本方差.

4. 连续产生服从标准正态分布的随机变量，直到产生 n 个为止，这里要求 $n \geqslant 100$ 且 $S/\sqrt{n} < 0.1$，其中 S 是这 n 个数值的样本标准差.

   (a) 你认为将产生多少个正态分布的随机变量？

   (b) 你已产生多少个正态分布的随机变量？

   (c) 所产生的正态分布随机变量的样本均值是什么？

   (d) 样本方差是什么？

   (e) 解释 (c), (d) 的结论，这些结论是否令人惊讶？

5. 除去考虑连续产生标准正态随机变量直到 $S/\sqrt{n} < 0.01$ 之外，重做习题 4.
6. 通过产生随机数估计 $\int_0^1 \exp(x^2)\mathrm{d}x$，至少产生 100 个值，直到估计的标准差小于 0.01 时，停止产生新的随机数.
7. 为估计 $E[X]$，我们已经得到 $X_1, \cdots, X_{16}$ 的模拟结果如下：10, 11, 10.5, 11.5, 14, 8, 13, 6, 15, 10, 11.5, 10.5, 12, 8, 16, 5. 根据这些数据，如果要使 $E[X]$ 的估计的标准差小于 0.1，大约需要再模拟多少个数据？

习题 8 和习题 9 与估计 e 有关.

133

8. 可以证明如果我们添加随机数直到它们的和超过 1，则添加数的个数的期望等于 e，也就是说，如果

$$N = \min\left\{n : \sum_{i=1}^{n} U_i > 1\right\}$$

则 $E[N] = \mathrm{e}$.

(a) 根据 1000 次模拟的结果，应用这个方法估计 e.

(b) 估计 (a) 中估计量的方差并给出 e 的置信度为 95% 的置信区间.

9. 考虑一个随机序列，用 M 表示第一个比它前面的元素小的那个元素. 也就是说，

$$M = \min\{n : U_1 \leqslant U_2 \leqslant \cdots \leqslant U_{n-1} > U_n\}$$

(a) 证明 $P\{M > n\} = \frac{1}{n!}$，$n \geqslant 0$.

(b) 应用等式 $E[M] = \sum_{n=0}^{\infty} P\{M > n\}$，证明 $E[M] = \mathrm{e}$.

(c) 根据 1000 次模拟的结果，应用 (b) 估计 e.

(d) 估计 (c) 中估计量的方差并给出 e 的置信度为 95% 的置信区间.

10. 应用第三章例 3a 的方法，要给出一个长度小于 0.1 的置信度为 95% 的 π 的置信区间，问需要模拟多少次？
11. 在区间长度不超过 0.01 的要求下，重做习题 10.
12. 为估计 θ，我们产生了 20 个均值为 θ 的独立的数值，分别是

102, 112, 131, 107, 114, 95, 133, 145, 139, 117
93, 111, 124, 122, 136, 141, 119, 122, 151, 143

如果要以 99% 的把握保证 θ 的最终估计的偏差在 ± 0.5 之内，你认为我们需要再产生多少个数值？

13. 设 $X_1, \cdots, X_n$ 是独立同分布的随机变量，均值 θ 未知，对给定的常数 $a < b$，我们对估计 $p = P\{a < \sum_{i=1}^{n} X_i/n - \mu < b\}$ 感兴趣.

(a) 说明怎样用自助法估计 p.

(b) 如果 $n = 10$ 并且 X_i 的值分别是 56, 101, 78, 67, 93, 87, 64, 72, 80, 69，估计 p（取 $a = -5, b = 5$）.

134

在下面的三个习题中，$X_1, \cdots, X_n$ 是一组来自方差 σ^2（未知）的分布的样本. 我们计划用样本方差 $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ 估计 σ^2，并想用自助法估计 $\mathrm{Var}(S^2)$.

14. 如果 $n = 2$，$X_1 = 1$ 和 $X_2 = 3$，$\mathrm{Var}(S^2)$ 的自助法估计是什么？
15. 如果 $n = 15$，数据是

$$5, 4, 9, 6, 21, 17, 11, 20, 7, 10, 21, 15, 13, 16, 8$$

则通过模拟求 $\mathrm{Var}(S^2)$ 的自助法估计的近似值.

16. 考虑一个顾客到达服从参数为 4.0 的泊松过程的单服务员系统. 当顾客到达时，只有系统中的顾客数小于或等于 3 时，他 (或她) 才进入系统. 顾客接受服务的时间服从参数为 4.2 的指数分布. 在 $T = 8$ 以后 (时间单位是小时)，不允许后来的顾客进入. 通过模拟估计进入系统的顾客在系统中平均花费的时间，并应用自助法，估计所给估计量的均方误差.

参考文献

[1] Bratley, P., B. L. Fox, and L. E. Schrage, *A Guide to Simulation*, 2nd ed. Springer-Verlag, New York, 1988.

[2] Crane, M. A., and A. J. Lemoine , *An Introduction to the Regenerative Method for Simulation Analysis.* Springer-Verlag, New York, 1977.

[3] Efron, B., and R. Tibshirani, *Introduction to the Bootstrap.* Chapman-Hall, New York, 1993.

[4] Kleijnen, J. P. C., *Statistical Techniques in Simulation*, Parts 1 and 2. Marcel Dekker, New York, 1974/1975.

[5] Law, A. M., and W. D. Kelton, *Simulation Modelling and Analysis*, 3rd ed. McGraw-Hill, New York, 1997.

135

第 8 章　方差缩减技术

引　言

在模拟研究中，一个典型的情况是：人们感兴趣的是想确定某量 θ，而 θ 是随机模型的一个参数. 为估计 θ，通过对模型进行模拟得到输出数据 X，它满足 $\theta = E[X]$. 进行重复模拟直到 n 次后停止，记第 i 次模拟输出的变量为 X_i，此时 θ 的估计由 $\bar{X} = \sum_{i=1}^{n} X_i/n$ 给出. 因为该估计是 θ 的无偏估计，所以它的均方误差就等于它的方差，即

$$\text{MSE} = E[(\bar{X} - \theta)^2] = \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n}$$

于是，如果我们能够得到比 $\bar{X}$ 的方差还小的 θ 的无偏估计，这样我们就可以对 θ 的估计进行改进.

在本章中，我们将尽力寻找各种不同的方法以减小模拟估计 $\bar{X}$ 的方差 (有时称 $\bar{X}$ 为原始估计).

在介绍方差缩减技术之前，我们先通过例子说明即使在十分简单的模型中，原始模拟估计也存在潜在的缺陷.

例 8a　质量控制　考虑一个逐个生产产品的过程，并假设这些产品的质量指标是可测量的. 当过程处于“可控”状态时，这些测量值 (适当标准化后) 来自标准正态分布. 进一步假设当过程处于“失控”状态时，这些值的分布从正态分布变成其他分布.

为了监测过程何时进入“失控”状态，指数加权移动平均控制图是一种经常使用的方法. 设 $X_1, X_2, \cdots$ 是一列观测值，对确定的值 α，$0 \leqslant \alpha \leqslant 1$，定义序列 S_n，如下 ($n \geqslant 0$)：

$$S_0 = 0$$

$$S_n = \alpha S_{n-1} + (1-\alpha) X_n, \quad n \geqslant 1$$

由于当过程处于可控状态时，所有的 X_n 的均值都是 0，因此，容易证明，在此条件下，指数加权移动平均值 S_n 的均值也是 0. 对于给定的 α 值，由指数加权移动平均控制图规则确定一个常数 B，当 $|S_n|$ 大于 B 时，认为过程“失控”. 也就是说，过程失控的时刻

$$N = \min\{n : |S_n| > B\}$$

是随机的.

很明显，即使过程仍在正常运行状态（也就是说即使产生的数据仍是服从标准正态分布的数据），$|S_n|$ 也有可能超过 B，即认为过程失控. 于是，为了确保这种情况不经常发生，需要慎重地选取 α 和 B 的值，使得当 $X_n, n \geqslant 1$ 确实是来自标准正态分布时，$E[N]$ 的值较大. 在过程可控的条件下，假设 800 作为 $E[N]$ 的值是可以接受的. 进一步假设有论断：如取 $\alpha = 0.9$ 和 $B = 0.8$，则可使 $E[N]$ 的值大致是 800. 我们如何检验这一论断？

检验上述结论的一个方法就是进行统计模拟. 即我们产生服从标准正态分布的随机变量 $X_n, n \geqslant 1$ 直到 $|S_n|$ 的值超过 0.8(在 $|S_n|$ 的定义中取 $\alpha = 0.9$). 如果用 N_1 表示直到 $|S_n|$ 的值超过 0.8 时已产生的服从正态分布的数据的个数，则对于第一次模拟运行，我们得到输出变量 N_1，然后接着进行模拟，所有模拟的输出数据的平均值就是 $E[N]$ 的估计.

然而，假设在过程可控的前提下，我们想以 99% 的把握保证 $E[N]$ 的估计精确到 ± 0.1 以内. 由于正态随机变量以 99% 的可能性落在偏离均值不超过 ± 2.58 个标准差的范围内（即 $Z_{0.005} = 2.58$），如果将需要运行的次数记作 n，则 n 需满足

$$\frac{2.58\sigma_n}{\sqrt{n}} \approx 0.1$$

138

其中 σ_n 是基于前 n 个数据的样本标准差. 现在，σ_n 将近似等于 N 的标准差 $\sigma(N)$，我们论证它近似等于 $E[N]$. 论证如下：由于假定过程仍处于可控状态，则绝大部分时间内指数加权移动平均值在原点附近移动，其绝对值偶尔有机会接近或大于 B. 运行一段时间后，或者移动平均的绝对值超过 B，过程运行结束，或者得到一串正态数据消除了移动平均值较大的情况. （实际中这后一种情况是可能的，因为 S_i 是由其以前值相继乘以 0.9 而累积起来的，因此以前值不再发挥作用）. 因此，如果我们知道在某一确定时间 k，过程仍没有失控，则无论 k 取何值，我们均有理由认为 S_k 的值在原点周围. 换句话说，从直观上看，移动平均值超出控制限的时间之分布基本上是无记忆的，即它近似是一个服从指数分布的随机变量. 又因为对于服从指数分布的随机变量 Y，有 $\mathrm{Var}(Y) = (E[Y])^2$，故当过程仍处于完全受控状态时，$\sigma(N) \approx E[N]$. 因此，若原先论断 $E[N] \approx 800$ 是正确的，则需要运行的次数是

$$\sqrt{n} \approx 25.8 \times 800$$

或者

$$n \approx (25.8 \times 800)^2 \approx 4.26 \times 10^8$$

另外，因为每次运行需要大约 800 个正态随机变量（再次假设论断的 $E[N] \approx 800$

是正确的)，我们得到做这个模拟将大约需要 $800 \times 4.26 \times 10^8 \approx 3.41 \times 10^{11}$ 个正态随机变量，这是一个可怕的工作. □

8.1 对偶变量的应用

假设我们对用模拟方法估计 $\theta = E[X]$ 感兴趣，且假设已产生均值是 θ 的同分布的随机变量 X_1 和 X_2，则

$$\mathrm{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}[\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)]$$

因此，如果 X_1 和 X_2 不是独立的，而是负相关的，则上式对于减小方差是有帮助的.

如何安排 X_1 和 X_2 使它们负相关呢？为此，我们假设 X_1 是 m 个随机数的函数，也就是假设

$$X_1 = h(U_1, U_2, \cdots, U_m)$$

139

其中 $U_1, \cdots, U_m$ 是 m 个相互独立的随机数. 如果 U 是一个随机数，即服从 $(0,1)$ 上的均匀分布，则 $1-U$ 也是一个随机数. 因此随机变量

$$X_2 = h(1-U_1, 1-U_2, \cdots, 1-U_m)$$

与 X_1 的分布相同. 另外，由于 $1-U$ 显然是与 U 负相关的，则我们希望得到 X_2 与 X_1 也是负相关的. 实际上，在 h 是其每一个自变量的单调（增或减）函数的情况下，可以证明该结论成立. [此结论实际上是两个单调增（或减）的独立随机变量的集函数是正相关这个一般结论的特殊情况，这两个结论的证明在本章的附录中给出.] 因此，在这种情况下，在利用 $U_1, \cdots, U_m$ 计算出 X_1 之后，我们可以利用 $1-U_1, 1-U_2, \cdots, 1-U_m$ 计算出 X_2，这将比使用新产生的 m 个随机数的独立的集合计算 X_2 更好. 并且，我们还可以收到事半功倍的效果：不但估计结果的方差较小（至少在 h 是单调函数时是如此)，而且可以节省生成第二个随机数集的时间. □

例 8b 模拟可靠性函数 考虑由 n 个部件组成的一个系统，每个部件都可能正常运转或失效. 令

$$s_i = \begin{cases} 1 & \text{如果部件 } i \text{ 正常运转} \\ 0 & \text{其他情况} \end{cases}$$

称 $\boldsymbol{s} = (s_1, \cdots, s_n)$ 是状态向量. 并假设存在一个非减函数 $\phi(s_1, \cdots, s_n)$ 满足

$$\phi(s_1, \cdots, s_n) = \begin{cases} 1 & \text{如果在状态向量}(s_1, \cdots, s_n)\text{下系统正常工作} \\ 0 & \text{其他情况} \end{cases}$$

此函数 $\phi(s_1,\cdots,s_n)$ 叫做结构函数.

一些常用的结构函数列举如下：

(a) **串联结构**：串联结构函数

$$\phi(s_1,\cdots,s_n)=\min_i s_i$$

串联结构系统正常工作当且仅当其中的每一个部件都正常运转.

(b) **并联结构**：并联结构函数

$$\phi(s_1,\cdots,s_n)=\max_i s_i$$

140
并联结构系统正常工作只需有不少于一个部件正常运转.

(c) **k/n 系统**：结构函数

$$\phi(s_1,\cdots,s_n)=\begin{cases}1 & \text{如果} \sum_{i=1}^n s_i \geqslant k\\ 0 & \text{其他情况}\end{cases}$$

叫做 k/n 结构函数. 由于 $\sum_{i=1}^n s_i$ 表示运转的部件的个数，k/n 系统正常工作要求 n 个部件中至少有 k 个部件正常运转.

应注意到串联系统是一个 n/n 系统，而并联结构系统是一个 $1/n$ 系统.

(d) **桥结构**：有五个部件组成的系统，若其结构函数

$$\phi(s_1,s_2,s_3,s_4,s_5)=\max(s_1s_3s_5,s_2s_3s_4,s_1s_4,s_2s_5)$$

则称其是一个桥结构. 这样的系统可以用图 8-1 粗略表示. 该示意图的想法是：如果一个信号能从左到右通过系统，则表明系统正常工作. 如果信号可以通过节点 i，则表明第 i 个部件正常运转，桥结构函数公式的证明留作习题.

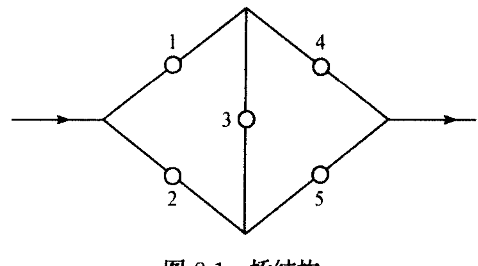


图 8-1 桥结构

现在我们假设部件所处的状态，记为 S_i, $i=1,\cdots,n$, 是相互独立的随机变量，且满足

$$P\{S_i=1\}=p_i=1-P\{S_i=0\},\ i=1,\cdots,n$$

令

$$\begin{aligned} r(p_1,\cdots,p_n) &= P\{\phi(S_1,\cdots,S_n)=1\} \\ &= E[\phi(S_1,\cdots,S_n)] \end{aligned}$$

函数 $r(p_1,\cdots,p_n)$ 叫做*可靠性函数*. 它表示第 i 个部件以概率 p_i 正常运转 $(i=1,\cdots,n)$ 且各部件相互独立时，系统正常工作的概率.

141

对于一个串联系统，

$$\begin{aligned} r(p_1,\cdots,p_n) &= P\{S_i=1, 对所有的 i=1,\cdots,n\} \\ &= \prod_{i=1}^{n} P\{S_i=1\} = \prod_{i=1}^{n} p_i \end{aligned}$$

对于一个并联系统，

$$\begin{aligned} r(p_1,\cdots,p_n) &= P\{S_i=1, \quad 至少有一个\ i, i=1,\cdots,n\} \\ &= 1-P\{S_i=0, \quad 对所有的\ i=1,\cdots,n\} \\ &= 1-\prod_{i=1}^{n} P\{S_i=0\} \\ &= 1-\prod_{i=1}^{n}(1-p_i) \end{aligned}$$

然而，对大多数系统而言，可靠性函数的计算仍是个难应对的问题.（即使像 5/10 系统或桥系统这样小的系统，计算也十分冗长.）假设对一个给定的非减结构函数 ϕ 和给定的概率 $p_1,\cdots,p_n$，我们感兴趣的是应用模拟方法估计

$$r(p_1,\cdots,p_n) = E[\phi(S_1,\cdots,S_n)]$$

现在我们可以通过产生服从均匀分布的随机变量 $U_1,\cdots,U_n$ 来模拟 S_i，并令

$$S_i = \begin{cases} 1 & 如果\ U_i < p_i \\ 0 & 其他情况 \end{cases}$$

于是我们看到

$$\phi(S_1,\cdots,S_m) = h(U_1,\cdots,U_n)$$

其中 h 是 $U_1,\cdots,U_n$ 的单调减函数. 因此

$$\mathrm{Cov}(h(\boldsymbol{U}), h(\boldsymbol{1}-\boldsymbol{U})) \leqslant 0$$

应用对偶变量方法，通过 $U_1, \cdots, U_n$ 生成 $h(U_1, \cdots, U_n)$ 和 $h(1-U_1, \cdots, 1-U_n)$ 得到的估计比用产生独立的随机数集生成第二个 h 的估计的方差小. □

142

通常情况下，模拟得到的相关输出结果是输入随机变量 $Y_1, \cdots, Y_m$ 的函数. 也就是说，相关输出结果可表示为 $X = h(Y_1, \cdots, Y_m)$. 假设 Y_i 的分布函数是 F_i, $i = 1, \cdots, m$. 如果这些输入的变量是通过逆变换方法产生的，X 可写成

$$X = h(F_1^{-1}(U_1), \cdots, F_m^{-1}(U_m))$$

其中 $U_1, \cdots, U_m$ 是独立的随机数. 如果分布函数是增函数，则它的逆函数也是增函数. 于是，如果 $h(y_1, \cdots, y_m)$ 是它的自变量的单调函数，则 $h(F_1^{-1}(U_1), \cdots, F_m^{-1}(U_m))$ 是它的自变量 U_i 的单调函数. 因此，应用对偶变量方法，通过 $U_1, \cdots, U_m$ 先算出 X_1，再利用 $1-U_1, \cdots, 1-U_m$ 算出 X_2 得到的估计的方差比用产生新随机数集算出 X_2 得到的估计的方差小. □

例 8c　模拟排队系统　考虑一个给定的排队系统，用 D_i 表示第 i 个到达的顾客在排队系统中等待的时间，假设我们感兴趣的是通过模拟这个系统估计 $\theta = E[X]$，其中

$$X = D_1 + \cdots + D_n$$

是前 n 个到达的顾客在排队系统中等待时间的总和，用 $I_1, \cdots, I_n$ 分别表示前 n 个到达者相邻两个之间的时间间隔 (即：I_j 表示第 $j-1$ 个与 j 个到达者之间的时间间隔)，并用 $S_1, \cdots, S_n$ 分别表示这个系统中前 n 个顾客接受服务的时间. 假设这些随机变量都是相互独立的. 在许多系统中，X 是 $2n$ 个随机变量 $I_1, \cdots, I_n, S_1, \cdots, S_n$ 的函数. 也就是说，

$$X = h(I_1, \cdots, I_n, S_1, \cdots, S_n).$$

对一个给定的顾客来说 (当然是在特定的模型下)，当其他的顾客接受服务的时间增加，通常他等待的时间也会增加，当相邻到达的顾客之间的时间间隔增加，通常他等待的时间就会减少. 对许多模型来说，h 是它的自变量的单调函数. 因此，如果用逆变换法产生随机变量 $I_1, \cdots, I_n, S_1, \cdots, S_n$，则应用对偶变量方法，得到的估计有较小的方差. 也就是说，如果开始用 $2n$ 个随机数 $U_i, i = 1, \cdots, 2n$ 通过变换 $I_i = F_i^{-1}(U_i), S_i = G_i^{-1}(U_{n+i})$ 产生到达的时间间隔和接受服务的时间，其中 F_i, G_i 分别是 I_i, S_i 的分布函数，则第二个模拟应该采用同样的方式进行，此时，应用随机数 $1-U_i, i = 1, \cdots, 2n$ 得到估计的方差比用重新产生 $2n$ 个随机数集得到的估计的方差要小. □

143

下例表明了应用对偶变量有时可得到的某种改进.

例 8d 假设我们感兴趣的是通过模拟估计

$$\theta = E[e^U] = \int_0^1 e^x dx$$

（当然，我们知道 $\theta = e - 1$，然而，该例子的目的是看应用对偶变量可能得到何种改进.）由于函数 $h(u) = e^u$ 显然是单调函数，则知应用对偶变量方法可以使方差减小，现在我们确定方差可以减小多少. 首先注意到

$$\begin{aligned}\mathrm{Cov}(e^U, e^{1-U}) &= E[e^U e^{1-U}] - E[e^U]E[e^{1-U}] \\ &= e - (e-1)^2 = -0.2342\end{aligned}$$

又因为

$$\begin{aligned}\mathrm{Var}(e^U) &= E[e^{2U}] - (E[e^U])^2 \\ &= \int_0^1 e^{2x} dx - (e-1)^2 \\ &= \frac{e^2-1}{2} - (e-1)^2 \\ &= 0.2420\end{aligned}$$

则由独立随机变量和的方差公式知

$$\mathrm{Var}\left(\frac{\exp\{U_1\} + \exp\{U_2\}}{2}\right) = \frac{\mathrm{Var}(e^U)}{2} = 0.1210$$

然而，应用对偶变量 U 和 $1-U$ 的结果的方差为

$$\mathrm{Var}\left(\frac{e^U + e^{1-U}}{2}\right) = \frac{\mathrm{Var}(e^U)}{2} + \frac{\mathrm{Cov}(e^U, e^{1-U})}{2} = 0.0039$$

该方差减少了 96.7%. □

例 8e 估计 e 考虑一个随机数序列 $U_i, i = 1, 2, \cdots$，定义 N 如下：

$$N = \min\{n : n \geqslant 2, U_n > U_{n-1}\}$$

144

则

$$P\{N > n\} = P\{U_1 \geqslant U_2 \geqslant \cdots \geqslant U_n\} = \frac{1}{n!}$$

最后一个等式是由于 $U_1, \cdots, U_n$ 的所有可能排序都是等可能的. 因此，

$$P\{N = n\} = P\{N > n-1\} - P\{N > n\} = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}$$

于是

$$E[N] = \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = \mathrm{e}$$

和

$$E[N^2] = \sum_{n=2}^{\infty} \frac{n}{(n-2)!} = \sum_{n=2}^{\infty} \frac{2}{(n-2)!} + \sum_{n=2}^{\infty} \frac{n-2}{(n-2)!}$$
$$= 2\mathrm{e} + \sum_{n=3}^{\infty} \frac{1}{(n-3)!} = 3\mathrm{e}$$

于是

$$\mathrm{Var}(N) = 3\mathrm{e} - \mathrm{e}^2 \approx 0.7658$$

因此，可通过产生随机数及利用随机数首次超过其前面的随机数的停时，对 e 进行估计.

如果我们使用对偶变量，则我们也可以令

$$M = \min\{n : n \geqslant 2, 1 - U_n > 1 - U_{n-1}\} = \min\{n : n \geqslant 2, U_n < U_{n-1}\}$$

由于在 N 和 M 的可能取值中，2 是其中的一个，而其他的均超过 2，尽管它们不是 U_n 的单调函数，但看起来估计量 $(N+M)/2$ 应该比两个独立的同 N 分布的随机变量的平均作为估计的方差要小. 在确定 $\mathrm{Var}(N+M)$ 之前，首先考虑一个有用的随机变量 N_a，它的分布与在给定 $U_2 \leqslant U_1$ 下，直到出现一个随机数大于它前面随机数时的随机数的个数的条件分布相同. 因此，我们有

145

$$N = 2, \qquad \text{以概率 } 1/2$$
$$N = 2 + N_a, \qquad \text{以概率 } 1/2$$

因此，

$$E[N] = 2 + \frac{1}{2}E[N_a]$$
$$E[N^2] = \frac{1}{2} \times 4 + \frac{1}{2}E[(2+N_a)^2]$$
$$= 4 + 2E[N_a] + \frac{1}{2}E[N_a^2]$$

于是，

$$\mathrm{Var}(N_a) = 14\mathrm{e} - 4\mathrm{e}^2 - 8 \approx 0.4997$$

现在考虑随机变量 N 和 M. 容易看出得到随机数前两个观测值以后, N 和 M 的取值中有一个等于 2, 其他的等于 2 加上一个与 N_a 同分布的随机变量. 因此,

$$\mathrm{Var}(N+M)=\mathrm{Var}(4+N_a)=\mathrm{Var}(N_a)$$

故有,

$$\frac{\mathrm{Var}(N_1+N_2)}{\mathrm{Var}(N+M)}\approx\frac{1.5316}{0.4997}\approx 3.065$$

于是, 应用对偶变量法使估计的方差大概为原来的 1/3. □

在正态随机变量的情况下, 若均值是 μ 方差是 σ^2, 则我们可以应用对偶变量法, 先产生这样的正态随机变量 Y, 然后将 $2\mu-Y$ 作为对偶变量, 它也是均值为 μ 方差为 σ^2 的正态随机变量, 且显然与 Y 负相关. 如果我们用模拟方法计算 $E[h(Y_1,\cdots,Y_n)]$, 其中 Y_i 是均值为 μ_i 的独立正态随机变量, $i=1,\cdots,n$, h 是一个单调函数, 则应用对偶方法, 先产生 n 个正态随机变量 $Y_1,\cdots,Y_n$ 并计算 $h(Y_1,\cdots,Y_n)$, 再应用对偶变量 $2\mu_i-Y_i, i=1,\cdots,n$ 计算 h 的下一个模拟值, 用此方法得到的估计的方差比再产生一组 n 个正态随机变量得到的估计的方差要小. □ 146

8.2 控制变量法的应用

再次假设我们想用模拟方法估计 $\theta=E[X]$, 这里 X 是模拟的输出结果. 对另外的模拟输出结果 Y, 期望值 $E[Y]=\mu_y$ 已知. 则对任意常数 c,

$$X+c(Y-\mu_y)$$

仍然是 θ 的无偏估计. 为了确定 c 的最优值, 注意到

$$\begin{aligned}\mathrm{Var}(X+c(Y-\mu_y))&=\mathrm{Var}(X+cY)\\&=\mathrm{Var}(X)+c^2\mathrm{Var}(Y)+2c\mathrm{Cov}(X,Y)\end{aligned}$$

通过简单计算可得, 在 $c=c^*$ 时上式达到最小值, 这里

$$c^*=-\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} \tag{8.1}$$

此时估计的方差是

$$\mathrm{Var}(X+c^*(Y-\mu_y))=\mathrm{Var}(X)-\frac{[\mathrm{Cov}(X,Y)]^2}{\mathrm{Var}(Y)} \tag{8.2}$$

数量 Y 叫做模拟估计 X 的控制变量. 为说明 Y 为什么会起到控制作用, 注意到当 X 与 Y 正（负）相关时, c^* 是负（正）的. 所以假设 X 与 Y 正相关, 其意思

粗略说就是，当 Y 较大时，X 也较大，反之也成立. 因此，如果模拟结果 Y 的值偏大（小），即大（小）于 Y 的已知均值 μ_y，则 X 有可能真的也大（小）于它的均值 θ，因此我们可以通过减小（增加）估计 X 的值对其进行修正，这个工作可以通过 c^* 取值为负（正）来完成. 当 X 与 Y 负相关时，可做类似的讨论.

将等式 (8.2) 两端同时除以 $\mathrm{Var}(X)$，得到

$$\frac{\mathrm{Var}(X + c^*(Y - \mu_y))}{\mathrm{Var}(X)} = 1 - \mathrm{Corr}^2(X, Y)$$

其中

147

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

是 X 与 Y 的相关系数. 因此，应用控制变量 Y 使估计的方差减少了 $100\mathrm{Corr}^2(X, Y)\%$.

一般情况下，事先并不知道 $\mathrm{Cov}(X, Y)$ 和 $\mathrm{Var}(Y)$ 的值，必须通过模拟对它们进行估计. 如果得到 n 次模拟的输出结果 $X_i, Y_i, i = 1, \cdots, n$，应用

$$\widehat{\mathrm{Cov}}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

和

$$\widehat{\mathrm{Var}}(Y) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

估计 $\mathrm{Cov}(X, Y)$ 和 $\mathrm{Var}(Y)$，我们可以得到 c^* 的近似 $\widehat{c}^*$，其中

$$\widehat{c}^* = -\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}.$$

控制估计量的方差

$$\mathrm{Var}(\bar{X} + c^*(\bar{Y} - \mu_y)) = \frac{1}{n}\left(\mathrm{Var}(X) - \frac{\mathrm{Cov}^2(X, Y)}{\mathrm{Var}(Y)}\right).$$

可以应用 $\mathrm{Cov}(X, Y)$ 的估计结合 $\mathrm{Var}(X)$ 和 $\mathrm{Var}(Y)$ 的估计进行估计.

注 另一种算法是利用简单线性回归模型的标准程序包进行计算. 如果我们考虑简单线性回归模型

$$X = a + bY + e$$

其中 e 是随机变量，其均值是 0，方差是 σ^2. $\widehat{a}$ 和 $\widehat{b}$ 分别是基于数据 $X_i, Y_i, i = 1, \cdots, n$ 的 a 和 b 的最小二乘估计，其中

$$\widehat{b} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$\widehat{a} = \bar{X} - \widehat{b}\bar{Y}$$

因此, $\widehat{b} = -\widehat{c}^*$. 并且, 由

$$\bar{X} + \widehat{c}^* \ (\bar{Y} - \mu_y) = \bar{X} - \widehat{b}\ (\bar{Y} - \mu_y) = \widehat{a} + \widehat{b}\mu_y$$

148

可得控制变量估计是估计回归线在点 $Y = \mu_y$ 的值. 又因为 σ^2 的回归估计 $\widehat{\sigma}^2$ 是 $\mathrm{Var}(X - \widehat{b}Y) = \mathrm{Var}(X + \widehat{c}^*Y)$ 的估计, 所以 $\widehat{\sigma}^2/n$ 是控制变量估计 $\bar{X} + \widehat{c}^*(\bar{Y} - \mu_y)$ 的方差的估计. □

例 8f 假设我们想应用模拟方法估计例 8b 中的可靠性函数

$$r(p_1, \cdots, p_n) = E[\phi(S_1, \cdots, S_n)]$$

其中

$$S_i = \begin{cases} 1 & \text{如果 } U_i < p_i \\ 0 & \text{其他情况} \end{cases}$$

由于 $E[S_i] = p_i$, 则可得

$$E\left[\sum_{i=1}^{n} S_i\right] = \sum_{i=1}^{n} p_i.$$

因此, 我们可以用运行部件数 $Y \equiv \sum_{i=1}^{n} S_i$ 作为估计 $X \equiv \phi(S_1, \cdots, S_n)$ 的控制变量, 由于 $\sum_{i=1}^{n} S_i$ 和 $\phi(S_1, \cdots, S_n)$ 都是 S_i 的增函数, 且二者是正相关的, 因此, c^* 的符号是负的. □

例 8g 考虑一个顾客到达服从强度函数为 $\lambda(s), s > 0$ 的非齐次泊松过程的排队系统. 假设每个顾客接受服务的时间是相互独立的随机变量, 有共同分布 G , 且与顾客到达的时间独立. 假设我们感兴趣的是估计在时刻 t 之前到达的所有顾客花费在系统中的总时间. 也就是说, 如果用 W_i 表示第 i 个到达的顾客花费在系统中的时间, 则我们感兴趣的是估计 $\theta = E[X]$, 其中

$$X = \sum_{i=1}^{N(t)} W_i$$

$N(t)$ 是时刻 t 之前到达的顾客数. 在这种情况下, 我们把所有这些顾客接受服务的总时间这个自然量作为控制变量, 也就是说, 若 S_i 表示第 i 顾客接受服务的时间,

则用

149

$$Y = \sum_{i=1}^{N(t)} S_i$$

作为控制变量，因为接受服务的时间与到达的顾客数 $N(t)$ 独立，可得

$$E[Y] = E[S]E[N(t)]$$

其中服务时间的均值 $E[S]$、到时刻 t 时平均到达的顾客数 $E[N(t)]$ 都是已知的. □

例 8h　假设在例 8d 中我们感兴趣的是用模拟方法计算 $\theta = E[\mathrm{e}^U]$. 这里用随机数 U 这个自然变量作为控制变量，看对原来的估计会有何种改进，注意到

$$\begin{aligned}\mathrm{Cov}(\mathrm{e}^U, U) &= E[U\mathrm{e}^U] - E[U]E[\mathrm{e}^U] \\ &= \int_0^1 x\mathrm{e}^x \mathrm{d}x - \frac{\mathrm{e}-1}{2} \\ &= 1 - \frac{\mathrm{e}-1}{2} = 0.14086\end{aligned}$$

因为 $\mathrm{Var}(U) = 1/12$，则根据等式 (8.2) 可得

$$\begin{aligned}\mathrm{Var}\left(\mathrm{e}^U + c^*\left(U - \frac{1}{2}\right)\right) &= \mathrm{Var}(\mathrm{e}^U) - 12 \times (0.14086)^2 \\ &= 0.2420 - 0.2380 = 0.0039\end{aligned}$$

这里应用了例 8d 的结果 $\mathrm{Var}(\mathrm{e}^U) = 0.2420$. 在这种情况下，应用控制变量 U 可使方差减少 98.4%. □

例 8i　表的重排问题　给定一个包含 n 个元素的集合，编号从 1 到 n，按一定次序排列. 在每个单位时间点上，按需要从中取出一个，元素 i 被需要的概率是 $p(i), \sum_{i=1}^n p(i) = 1$. 取出后，放回表中，但不一定放回原来的位置. 例如，一般的重排规则是将需要的元素与它前面直接相邻的元素进行交换. 因此，如果 $n = 4$，并且当前的排列是 $1, 4, 2, 3$，则按照规则，若需要的元素是 2，则重排后结果为 $1, 2, 4, 3$. 从一个给定的排列开始，该排列等可能地为 $n!$ 种可能排列中的任意一个，应用这种互换规则，假设我们感兴趣的是确定前 N 个被需要的元素所在的位置和的期望. 怎样通过模拟方法高效率地完成这个任务？

下面给出一个有效的方法. 模拟上面问题的“自然”的方法是产生 $1, \cdots, n$ 的一个随机排列作为最初的排序，在接下来的 N 次交换的每次交换中，通过产生随机数 U 确定需交换的元素，如果 $\sum_{k=1}^{j-1} p(k) < U \leqslant \sum_{k=1}^{j} p(k)$，则对元素 j 进行交换. 然而，一个更好的方法是用随机数 U 的较小的值对应位置靠前的元素的方法产生所需要的元素. 也就是说，如果当前的排列是 $i_1, i_2, \cdots, i_n$，通过产生随机数

150

U，产生需要的元素，如果 $\sum_{k=1}^{j-1} p(i_k) < U \leqslant \sum_{k=1}^{j} p(i_k)$，则为 i_j 选取交换. 例如，如果 $n=4$ 并且当前的排列是 $3,1,2,4$，我们应该产生 U，如果 $U \leqslant p(3)$，则为 3 选取交换，如果 $p(3) < U \leqslant p(3)+p(1)$，则为 1 选取交换，依此类推. 由于随机数 U 的较小的值对应位置靠前的元素，我们可以用 $\sum_{r=1}^{N} U_r$ 作为控制变量，U_r 是在一次交换中为选取第 r 个需要的元素利用的随机数. 也就是说，如果 P_r 是在一次交换中选取的第 r 个需要的元素的位置. 不用原始估计 $\sum_{r=1}^{N} P_r$，而用

$$\sum_{r=1}^{N} P_r + c^* \left(\sum_{r=1}^{N} U_r - \frac{N}{2} \right)$$

作为估计，其中

$$c^* = -\frac{\mathrm{Cov}\left(\sum\limits_{r=1}^{N} P_r, \sum\limits_{r=1}^{N} U_r \right)}{\frac{N}{12}}$$

其中的协方差要用所有模拟数据进行估计.

虽然方差缩减依赖概率 $p(i), i=1,\cdots,n$ 和 N 的值，但是小规模的研究表明，当 $n=50$ 且 $p(i)$ 近似相等时，对 $15 \leqslant N \leqslant 50$ 的情况，控制估计的方差小于原估计方差的 1/2400. □

当然，也可以用多个变量做控制变量. 例如，如果模拟输出变量是 $Y_i, i=1,\cdots,k$，并且 $E[Y_i]=\mu_i$ 已知，则对任意常数 $c_i, i=1,\cdots,k$ 我们可以用

$$X + \sum_{i=1}^{k} c_i (Y_i - \mu_i)$$

作为 $E[X]$ 的无偏估计.

例 8j　21 点游戏　21 点游戏的玩法通常是发牌者洗多副牌，将用过的放一边，当剩余的牌数低于某个限定的数时，重新洗牌. 也就是说，每当新开一局时，发牌者都重新洗牌，我们感兴趣的是用模拟方法估计一个玩家每局获胜的期望 $E[X]$，[151] 假设玩家采取一些固定的策略，该策略可能是已经玩过的计点数的玩法，根据“点数”赌不同的数值. 我们假设该游戏是单个玩家对发牌者（庄家）.

这种游戏的随机性来自发牌者洗牌. 如果发牌者用 k 副 52 张的牌，我们可以通过产生从 1 到 $52k$ 的一个随机排列，生成一副洗好的牌；用 $I_1,\cdots,I_{52k}$ 表示这个排列. 如果令

$$u_j = I_j \bmod 13 + 1$$

和

$$v_j = \min(u_j, 10)$$

则 $v_j, j=1,\cdots,52k$ 表示洗好牌的点数，1 表示“A”.

用 N 表示一局中玩的手数，B_j 表示第 j 手赌注数. 为了减小方差，当玩家手中的牌比庄家好时，我们用较大的控制变量；当情况相反时，用较小的控制变量. 由于 19 点或更大的点是好的，我们定义

$$W_j = \begin{cases} 1 & \text{如果玩家第}j\text{手的两张牌的点数加在一起至少是 19} \\ 0 & \text{其他情况} \end{cases}$$

同样地令

$$Z_j = \begin{cases} 1 & \text{如果庄家第 }j\text{手的两张牌的点数加在一起至少是 19} \\ 0 & \text{其他情况} \end{cases}$$

因为 W_j 和 Z_j 分布相同，所以 $E[W_j - Z_j] = 0$，并且不难证明

$$E\left[\sum_{j=1}^{N} B_j(W_j - Z_j)\right] = 0.$$

因此，我们建议用 $\sum_{j=1}^{N} B_j(W_j - Z_j)$ 作为控制变量. 当然，并不清楚 19 点是否为最优值，应该试一试用 18 点或 20 点做临界值看结果如何. 然而，一些初步的研究结果表明 19 点是最优值，并且根据玩家的策略证明可使方差减小至少 15%. 如果我们用两个控制变量，方差会减小更多. 第一个控制变量定义如上，但不包括当这一手是 19 点或 20 点时，W_j 和 Z_j 为 1 的情况. 第二个控制变量可类似定义，但当这一手是 21 点时，它对应的是 1. □

152

当应用多个控制变量时，可以应用多元线性回归模型

$$X = a + \sum_{i=1}^{k} b_i Y_i + e$$

的计算机程序完成计算，其中 e 是随机变量，均值是 0，方差是 σ^2. 如果用 $\widehat{c}_i^*$ 表示最优值 c_i 的估计，$i = 1, \cdots, k$，则

$$\widehat{c}_i^* = -\widehat{b}_i, \qquad i = 1, \cdots, k$$

其中 $\widehat{b}_i$ 是 $b_i, i = 1, \cdots, k$ 的最小二乘回归估计. 控制估计的值由下式可得

$$\bar{X} + \sum_{i=1}^{k} \widehat{c}_i^*(\bar{Y}_i - \mu_i) = \widehat{a} + \sum_{i=1}^{k} \widehat{b}_i \mu_i.$$

也就是说，控制估计正好是多元回归线在点 $(\mu_1, \cdots, \mu_k)$ 的估计值.

控制估计的方差可由 σ^2 的回归估计除以模拟次数得到.

注 1 因为事先并不知道控制估计的方差，所以模拟通常需要分两步完成. 第一步通过少量的模拟给出 $\text{Var}(X + c^*(Y - \mu_y))$ 的粗略的估计. (应用 σ^2 的估计，通

过简单的线性回归程序可得到这个估计，这里 Y 和 X 是相互独立的变量.) 然后，在第二步模拟中通过确定模拟次数使最后估计的方差限定在可接受的范围内.

注 2 要说明控制变量方法的作用需将其与 θ 的估计联合起来考虑. 也就是说，假设 X 和 W 的值由模拟确定，并且假设 $E[X] = E[W] = \theta$. 则我们可以考虑如下形式的无偏估计

$$\alpha X + (1-\alpha)W$$

通过选取 α 使方差达到最小, 可得当 $\alpha = \alpha^*$ 时，这样的估计是最优估计，其中

$$\alpha^* = \frac{\mathrm{Var}(W) - \mathrm{Cov}(X, W)}{\mathrm{Var}(X) + \mathrm{Var}(W) - 2\mathrm{Cov}(X, W)} \tag{8.3}$$

153

如果 $E[Y] = \mu_y$ 已知，我们得到两个无偏估计 X 和 $X + Y - \mu_y$，联合估计可写成如下形式

$$(1-c)X + c(X + Y - \mu_y) = X + c(Y - \mu_y)$$

为了得到控制变量和联合估计的其他等价形式，假设 $E[X] = E[W] = \theta$. 则如果用均值已知的变量 $Y = X - W$，其均值为 0，作为控制变量，用 X 可得到如下形式的估计

$$X + c(X - W) = (1+c)X - cW$$

这是一个 $\alpha = 1 + c$ 的联合估计.

注 3 根据注 2 的解释，对偶变量法可以看作控制变量法的特殊情况. 也就是说，如果 $E[X] = \theta$，其中 $X = h(U_1, \cdots, U_n)$，则也有 $E[W] = \theta$，这里 $W = h(1 - U_1, \cdots, 1 - U_n)$. 因此，我们可以得到 $\alpha X + (1-\alpha)W$ 形式的联合估计. 因为 X 与 W 分布相同，$\mathrm{Var}(X) = \mathrm{Var}(W)$，所以可根据等式 (8.3) 得到 α 的最优值是 $\alpha = 1/2$，这正是对偶变量估计.

注 4 注 3 说明了为什么对偶变量通常不可能有效地与控制变量联合. 如果控制变量 Y 与 $h(U_1, \cdots, U_n)$ 有大的正（负）相关性，则它可能与 $h(1-U_1, \cdots, 1-U_n)$ 有大的负（正）相关. 因此，它不可能与对偶变量 $\frac{h(U_1,\cdots,U_n)+h(1-U_1,\cdots,1-U_n)}{2}$ 有大的相关性. □

8.3 缩减方差的条件期望法

回顾在第 2 章 2.10 节证明的条件方差公式

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

因为方差总是非负的，所以上式右端两项都是非负的，可得

$$\mathrm{Var}(X) \geqslant \mathrm{Var}(E[X|Y]) \tag{8.4}$$

假设现在我们感兴趣的是通过模拟研究确定 $\theta = E[X]$ 的值，其中 X 是模拟运行输出的变量. 又假设有第二个变量 Y，因而 $E[X|Y]$ 已知并且其值可以通过模拟确定. 由于

154

$$E[E[X|Y]] = E[X] = \theta$$

可知 $E[X|Y]$ 也是 θ 的一个无偏估计；从 (8.4) 可知 $E[X|Y]$ 作为 θ 的估计，优于 (原) 估计 X.

注　为了说明条件期望估计优于原估计，首先注意到我们是通过模拟来估计 θ 的未知值. 现在我们可以想象将模拟分两阶段进行：我们先观测随机变量 Y 的模拟值，然后再观测 X 的模拟值. 然而，如果得到 Y 的观测值后，可以计算 X 的 (条件) 期望值，则我们可以根据这个值给出 $E[X]$ 的一个估计，这个估计消除了模拟 X 的实际值增加的方差. □

就这一点来说，也许会考虑用 $\alpha X + (1-\alpha)E[X|Y]$ 类型的估计做进一步的改进. 然而，由等式 (8.3) 知这种类型的估计在 $\alpha = \alpha^*$ 时是最优的，其中

$$\alpha^* = \frac{\mathrm{Var}(E[X|Y]) - \mathrm{Cov}(X, E[X|Y])}{\mathrm{Var}(X) + \mathrm{Var}(E[X|Y]) - 2\mathrm{Cov}(X, E[X|Y])}$$

我们证明 $\alpha^* = 0$，以表明联合估计量 X 与 $E[X|Y]$ 并没有得到什么改进.

首先注意到

$$\begin{aligned}\mathrm{Var}(E[X|Y]) &= E[(E[X|Y])^2] - (E[E[X|Y]])^2 \\ &= E[(E[X|Y])^2] - (E[X])^2\end{aligned} \tag{8.5}$$

另一方面，

$$\begin{aligned}\mathrm{Cov}(X, E[X|Y]) &= E[XE[X|Y]] - E[X]E[E[X|Y]] \\ &= E[XE[X|Y]] - (E[X])^2 \\ &= E[E[XE[X|Y]|Y]] - (E[X])^2 \quad \text{将}Y\text{作为条件} \\ &= E[E[X|Y]E[X|Y]] - (E[X])^2 \quad \text{因为给定 } Y, E[X|Y]\text{是常数} \\ &= \mathrm{Var}(E[X|Y]) \quad \text{根据 (8.5)}\end{aligned}$$

因此，我们看到用 X 与 $E[X|Y]$ 的联合估计并没有使方差有额外的缩减.

155

我们通过一系列的例子说明“条件期望”的用法.

例 8k 让我们再次考虑用模拟方法估计 π. 在第 3 章的例 3a 中，我们给出了如何通过在面积为 4 中心在原点的正方形内均匀分布的随机点落在其半径为 1 的内接圆内的点的频率来估计 π. 特别地，令 $V_i = 2U_i - 1$，其中 $U_i, i = 1,2$ 是随机数，并令

$$I = \begin{cases} 1 & \text{如果 } V_1^2 + V_2^2 \leqslant 1 \\ 0 & \text{其他情况} \end{cases}$$

如同在例 3a 中一样，可得 $E[I] = \pi/4$.

可以通过应用 $E[I|V_1]$ 对 $\pi/4$ 的原始估计 I 进行改进. 因为

$$\begin{aligned} E[I|V_1 = v] &= P\{V_1^2 + V_2^2 \leqslant 1 | V_1 = v\} \\ &= P\{v^2 + V_2^2 \leqslant 1 | V_1 = v\} \\ &= P\{V_2^2 \leqslant 1 - v^2\} \qquad \text{由 } V_1 \text{和 } V_2 \text{ 独立} \\ &= P\{-(1-v^2)^{1/2} \leqslant V_2 \leqslant (1-v^2)^{1/2}\} \\ &= \int_{-(1-v^2)^{1/2}}^{(1-v^2)^{1/2}} \left(\frac{1}{2}\right) \mathrm{d}x \qquad \text{因为 } V_2 \text{是 } (-1,1) \text{上的均匀分布} \\ &= (1-v^2)^{1/2} \end{aligned}$$

所以，

$$E[I|V_1] = (1 - V_1^2)^{1/2}$$

并且估计量 $(1-V_1^2)^{1/2}$ 的均值也是 $\pi/4$，但方差比 I 的小. 因为

$$\begin{aligned} E\left[\left(1 - V_1^2\right)^{1/2}\right] &= \int_{-1}^{1} (1-x^2)^{1/2} \left(\frac{1}{2}\right) \mathrm{d}x \\ &= \int_0^1 (1-x^2)^{1/2} \mathrm{d}x \\ &= E[(1-U^2)^{1/2}] \end{aligned}$$

所以可以用估计 $(1-U^2)^{1/2}$ 进行一些简化，其中 U 是随机数.

容易得到估计 $(1-U^2)^{1/2}$ 在方差上的改进.

$$\begin{aligned} \mathrm{Var}[(1-U^2)^{1/2}] &= E[1-U^2] - \left(\frac{\pi}{4}\right)^2 \\ &= \frac{2}{3} - \left(\frac{\pi}{4}\right)^2 \approx 0.0498 \end{aligned}$$

156

在上述第一个等式中我们利用了 $\mathrm{Var}(W) = E[W^2] - (E[W])^2$. 另一方面，由 I 是服从二项分布的随机变量且均值为 $\pi/4$，可得

$$\mathrm{Var}(I) = \left(\frac{\pi}{4}\right)\left(1 - \frac{\pi}{4}\right) \approx 0.1686$$

这表明条件期望法使方差减少了 70.44%.（另外，对每次模拟也只需产生一个而不是两个随机数，尽管必须以计算平方根为代价.）

显然，因为函数 $(1-u^2)^{1/2}$ 是区间 $(0,1)$ 上的单调减函数，所以应用对偶变量可以对估计量 $(1-U^2)^{1/2}$ 进行改进. 也就是说，估计量

$$\frac{1}{2}[(1-U^2)^{1/2}+(1-(1-U)^2)^{1/2}]$$

的方差比估计量 $\frac{1}{2}[(1-U_1^2)^{1/2}+(1-U_2^2)^{1/2}]$ 的方差小.

另一种方法是应用控制变量改进估计量 $(1-U^2)^{1/2}$. 这种情况下，U^2 是一个自然的控制变量，由 $E[U^2]=1/3$，我们可以用

$$(1-U^2)^{1/2}+c\left(U^2-\frac{1}{3}\right)$$

形式的估计. 通过对协方差项的模拟估计，可以得到 c 的最优值，即 $c^*=-\text{Cov}[(1-U^2)^{1/2},U^2]/\text{Var}(U^2)$ 的估计.（我们也可以试着用 U 做控制变量；但这会与上面的有所不同，因为两个随机变量的相关性仅是它们"线性相关"的度量，而不是总体相关的度量. 但应用 U^2 作控制变量会有更好的改进；见习题 15.）□

例 81　假设 Y 是均值为 1 的指数随机变量，并假设在 $Y=y$ 的条件下，X 是正态随机变量，均值是 y，方差是 4. 如何用模拟方法有效地估计 $\theta=P\{X>1\}$? 原始的模拟方法是从产生 Y 的值开始，就是产生随机数 U，并令 $Y=-\ln(U)$；如果 $Y=y$，则产生均值为 y 方差为 4 的正态随机变量 X 的值，并令

157

$$I=\begin{cases}1 & \text{如果 } X>1\\ 0 & \text{如果 } X\leqslant 1\end{cases}$$

多次运行的 I 的平均值就是原始的模拟估计.

然而注意到，当 $Y=y$ 时，

$$Z=\frac{X-y}{2}$$

是一个标准正态随机变量. 于是，

$$E[I|Y=y]=P\{X>1|Y=y\}=P\left\{Z>\frac{1-y}{2}\right\}=\bar{\Phi}\left(\frac{1-y}{2}\right) \tag{8.6}$$

其中 $\bar{\Phi}(x)=1-\Phi(x)$ 是一个标准正态随机变量大于 x 的概率. 因此，多次运行得到的 $\bar{\Phi}(\frac{1-Y}{2})$ 的平均值将优于原始的模拟估计.

因为条件期望估计 $\bar{\Phi}(\frac{1-Y}{2})$ 是 Y 的单调函数，则应用对偶变量可进一步地改进估计的方差. 也就是说，可以用随机数 U 给出估计

$$\frac{\bar{\Phi}\left(\frac{1+\ln(U)}{2}\right)+\bar{\Phi}\left(\frac{1+\ln(1-U)}{2}\right)}{2}$$

除了应用对偶变量，还有一种可能是用 Y 做控制变量. 可是用对偶变量还是用控制变量，哪一个更有效只能通过模拟研究才能确定. □

在下面的例子中，我们用条件期望法有效地估计出了复合随机变量大于某一固定值的概率.

例 8m 设 $X_1, X_2, \cdots$ 是一个独立同分布的正的随机变量序列，并且与非负整值随机变量 N 独立. 随机变量

$$S = \sum_{i=1}^{N} X_i$$

被称做复合随机变量. 在保险应用中，X_i 表示第 i 个向保险公司索赔的数额，N 表示到规定时间 t 之前索赔的个数；S 是到时刻 t 为止索赔额的总数. 在这样的应用中，通常假设 N 是泊松随机变量（在这种情况下，S 被称做复合泊松随机变量）或者是混合泊松随机变量，我们说 N 是混合泊松随机变量是指如果存在另一个随机变量 Λ，在给定 $\Lambda = \lambda$ 时，N 的条件分布是均值为 λ 的泊松分布. 例如，如果 Λ 有概率密度函数 $g(\lambda)$，则混合泊松随机变量 N 的概率分布函数是 158

$$P\{N = n\} = \int_0^\infty \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!} g(\lambda)\mathrm{d}\lambda$$

当在某感兴趣的时间周期内发生的（泊松）事件数的均值由一个随机确定的“环境状态”所确定时，就出现了混合泊松随机变量. Λ 的分布函数被称做混合分布.

对某个确定的正常数 c，设想模拟估计

$$p = P\left\{\sum_{i=1}^{N} X_i > c\right\}$$

最初的估计方法是先产生 N 的值，即 $N = n$，然后产生 $X_1, \cdots, X_n$ 的值，应用它们确定最初模拟估计的值

$$I = \begin{cases} 1 & \text{当 } \sum_{i=1}^{N} X_i > c \\ 0 & \text{其他情况} \end{cases}$$

在多次模拟运行下，得到的 I 的平均值就是 p 的估计.

我们可以应用条件期望法改进前面的估计，从产生序列 X_i 的值开始，到产生值的和超过 c 为止. 用 M 表示所需要的运行次数，也就是

$$M = \min\left(n : \sum_{i=1}^{n} X_i > c\right)$$

如果产生的 M 的值是 m，则我们用 $P\{N \geqslant m\}$ 作为这次运行中 p 的估计. 可见这个估计的方差比最初模拟估计 I 的方差小. 其原因是 X_i 是正的，且

$$I = 1 \Longleftrightarrow N \geqslant M$$

因此，

$$E[I|M] = P\{N \geqslant M|M\}$$

又因为

159

$$P\{N \geqslant M|M = m\} = P\{N \geqslant m|M = m\} = P\{N \geqslant m\}$$

最后一个等式成立是因为 N 与 M 相互独立. 于是可知，若模拟得到的 M 的值是 $M = m$，则得到的 $E[I|M]$ 的值是 $P\{N \geqslant m\}$.

应用控制变量可以对上述的条件期望估计做进一步改进. 令 $\mu = E[X_i]$，并且定义

$$Y = \sum_{i=1}^{M}(X_i - \mu)$$

可以证明 $E[Y] = 0$. 为了直观上说明 Y 和条件期望估计 $P\{N \geqslant M|M\}$ 强相关，首先注意到当 m 大时，条件期望估计将小. 但是，因为 M 是所需要的加起来的总和超过 c 的 X_i 的个数，可知当 X_i 值小时，M 的值将大，这将使 Y 的值小. 也就是说，$E[I|M]$ 和 Y 的值同时趋向于变小. 类似讨论可表明：当 $E[I|M]$ 的值趋向于大时，Y 的值也趋向于大. 由此，显然可得 $E[I|M]$ 和 Y 强正相关，即 Y 是一个有效的控制变量. □

例 8n　有限容量的排队模型　考虑如下的排队系统：只有当系统中排队的顾客数少于 N 时才允许新到达的顾客进入该排队系统，而当新顾客到达时恰有 N 个顾客在系统中，则认为该顾客从系统流失. 进一步假设潜在的顾客以参数为 λ 的泊松过程到达；并假设我们感兴趣的是用模拟方法估计到某一确定的时间 t 时流失顾客数的期望.

每次上述系统的模拟都将到时间 t 为止. 对给定的一次模拟，用 L 表示流失的顾客数，则所有模拟中 L 的平均值是要求的数值 $E[L]$ 的（最初）模拟估计，以系统处于满容量的时间数作为条件，我们可以改进这个估计. 也就是说，我们不是用 L，而是用到时间 t 实际流失的顾客数来改进这个估计. 我们考虑 $E[L|T_C]$，其中 T_C 是在区间 $(0,t)$ 内恰好有 N 个顾客在系统中的总时间. 因为无论系统内发生什么，顾客总是以参数为 λ 的泊松过程到达，于是

$$E[L|T_C] = \lambda T_C$$

因此, 可以通过确定每次模拟中, 恰有 N 个顾客在系统中的总时间来改进上述估计. 若用 $T_{C,i}$ 表示第 i 次模拟时满容量的总时间, 则 $E[L]$ 的改进估计是 $\lambda\sum_{i=1}^{k}T_{C,i}/k$, 其中 k 是模拟的次数.（在效果上, 由于在给定满容量的时间是 T_C 时, 流失的顾客数的期望正好是 λT_C, 这个估计所做的是用实际条件期望而不是模拟一个有此均值的泊松随机变量, 因为它会增加估计的方差.）

160

如果到达过程是非齐次泊松过程, 强度函数是 $\lambda(s), 0\leqslant s\leqslant t$, 且只给满容量的总时间, 则我们无法计算出流失的顾客数的条件期望. 现在我们需要的是系统处于满容量的实际时间, 即系统何时处于满容量. 因此, 我们以系统处于满容量的区间为条件, 用 N_C 表示在 $(0,t)$ 内系统处于满容量的区间数, 并将它们记作 $I_1,\cdots,I_{N_C}$, 则

$$E[L|N_C,I_1,\cdots,I_{N_C}]=\sum_{i=1}^{N_C}\int_{I_i}\lambda(s)\mathrm{d}s$$

取上述模拟得到的平均值, 则得到 $E[L]$ 的一个估计, 且在均方误差的意义下, 该估计比用每次模拟流失数的平均值得到的模拟估计要好.

在估计 $E[L]$ 时, 也可以将上面的方法与其他方差缩减的技术结合起来使用. 例如, 如果我们用 M 表示到时间 t 时实际进入服务系统的顾客数, 如果 $N(t)$ 等于到时间 t 时到达的顾客数, 则

$$N(t)=M+L$$

求期望得到

$$\int_0^t\lambda(s)\mathrm{d}s=E[M]+E[L]$$

因此, $\int_0^t\lambda(s)\mathrm{d}s-M$ 也是 $E[L]$ 的无偏估计, 这建议我们用联合估计

$$\alpha\sum_{i=1}^{N_C}\int_{I_i}\lambda(s)\mathrm{d}s+(1-\alpha)\left(\int_0^t\lambda(s)\mathrm{d}s-M\right)$$

其中 α 的值由等式 (8.3) 给出或通过模拟估计得到. □

例 8o 假设我们想估计在一个排队系统中前 n 个顾客花费在系统中的总时间的期望. 也就是说, 如果 W_i 是第 i 个顾客在系统中花费的时间, 则我们感兴趣的是估计

$$\theta=E\left[\sum_{i=1}^{n}W_i\right]$$

161

用 S_i 表示第 i 个顾客到达时“系统的状态”, 并考虑估计

$$\sum_{i=1}^{n}E[W_i|S_i]$$

因为

$$E\left[\sum_{i=1}^{n} E[W_i|S_i]\right] = \sum_{i=1}^{n} E[E[W_i|S_i]] = \sum_{i=1}^{n} E[W_i] = \theta$$

所以此估计量是 θ 的一个无偏估计. 可以证明 [1], 在一大类的模型中, 这个估计的方差比最初的模拟估计 $\sum_{i=1}^{n} W_i$ 的方差小. (应该注意到, 尽管很显然 $E[W_i|S_i]$ 的方差比 W_i 的方差小, 由于协方差项的缘故, 这并不能推出 $\sum_{i=1}^{n} E[W_i|S_i]$ 的方差比 $\sum_{i=1}^{n} W_i$ 的方差小.)

假设数量 S_i, 即第 i 个顾客到达时系统所处的状态, 是能够使我们计算顾客花费在系统中的时间的条件期望的最小信息量. 例如, 假设仅有一名服务员, 且服务时间都服从均值是 μ 的指数分布, 则 S_i 将为 N_i, 即第 i 个顾客到达时系统中的顾客数. 在这种情况下,

$$E[W_i|S_i] = E[W_i|N_i] = (N_i + 1)\mu$$

这个结论是因为第 i 个到达的顾客将不得不等待 N_i 个服务时间 (一个服务时间是指第 i 个顾客到达时正在接受服务的顾客服务完成的时间, 但是, 由指数分布的无记忆性, 剩余的时间仍服从均值是 μ 的指数分布), 它们都有均值 μ, 并且加上自己的服务时间. 因此, 所有模拟运行得到的数量 $\sum_{i=1}^{n}(N_i + 1)\mu$ 的平均值是比 $\sum_{i=1}^{n} W_i$ 的平均值更好的估计. □

下面这个例子涉及到, 在靶形细胞集合被杀死之前, 没有被意外杀死的非靶形细胞个数的分布.

例 8p　考虑一个有 $n+m$ 个细胞的集合, 第 i 个细胞的权重为 ω_i, $i = 1,\cdots,n+m$. 假设第 $1,\cdots,n$ 个细胞是癌细胞, 而第 $n+1,\cdots,n+m$ 个细胞是正常的, 并假设以如下方式一次杀死一个细胞: 在任何时候, 如果 S 是当前仍活着的细胞的集合 (其他不在 S 中的细胞已被杀死), 下一个被杀死细胞是 i $(i \in S)$ 的概率为 $\frac{\omega_i}{\sum_{j\in S}\omega_j}$, 且与已被杀死细胞的次序独立. 因此, 细胞 i 以概率 $\frac{\omega_i}{\sum_{j=1}^{n+m}\omega_j}$ 第一个被杀死; 在给定第一个杀死的细胞是 i 的条件下, 下一个被杀死细胞是 k $(k \neq i)$ 的概率为 $\frac{\omega_k}{\sum_{j\neq i}\omega_j}$, 以此类推. 该过程持续到前 n 个细胞 (癌细胞) 都被杀死. 以 N 表示到癌细胞都被杀死时刻, 仍活着的正常细胞的个数. 我们感兴趣的是确定概率 $P\{N \geqslant k\}$.

[162]

在试图得到有效的模拟程序之前, 我们先考虑一个相关的模型, 在该模型中, 细胞 i, $i = 1,\cdots,n+m$ 在随机时间 T_i 被杀死, 其中 $T_1,\cdots,T_{n+m}$ 是相互独立的参数分别为 $\omega_i,\cdots,\omega_{n+m}$ 的指数随机变量. 由指数随机变量的无记忆性可知, 如

1. S. M. Ross, *Simulating Average Delay—Variance Reduction by Conditioning*, *Probability Eng. Informational Sci.* **2**(3),1988.

果 S 是当前活着的细胞的集合，如同在原始模型中一样，下一个将被杀死细胞是 i $(i \in S)$ 的概率为 $\frac{\omega_i}{\sum_{j\in S}\omega_j}$，这说明，在这个相关模型中，细胞被杀死次序的概率分布与原始模型中的概率分布相同. 以 N 表示细胞 $1,\cdots,n$ 都被杀死时刻仍活着的细胞个数. 如果以 $T^{(k)}$ 表示 $T_{n+1},\cdots,T_{n+m}$ 中的第 k 个最大值，则 $T^{(k)}$ 表示活着的正常细胞个数首次少于 k. 因此，为了使 N 的值达到最小值 k，到时刻 $T^{(k)}$ 所有的癌细胞必须已经被杀死. 也就是，

$$P\{N \geqslant k\} = P\{\max_{i\leqslant n} T_i < T^{(k)}\}$$

因此，

$$\begin{aligned}P\{N \geqslant k|T^{(k)}\} &= P\{\max_{i\leqslant n} T_i < T^{(k)}|T^{(k)}\}\\ &= \prod_{i=1}^{n}(1-\mathrm{e}^{-\omega_i T^{(k)}})\end{aligned}$$

其中最后一个等式应用了 T_i 的独立性. 因此，通过产生 m 个服从指数分布的随机变量 $T_{n+1},\cdots,T_{n+m}$，得到 $P\{N \geqslant k\}$ 的一个无偏的条件期望的估计. 以 $T^{(k)}$ 表示 $T_{n+1},\cdots,T_{n+m}$ 中的第 k 个最大值，得到估计 $\prod_{i=1}^{n}(1-\mathrm{e}^{-\omega_i T^{(k)}})$. 因为这个估计是产生的 $T_{n+1},\cdots,T_{n+m}$ 的增函数，如果用逆变换方法得到 T_i，可以进一步缩减方差. 因为这个估计将是所用的 m 个随机数的增函数，表明使用相反的变量将可以进一步缩减方差. 综合考虑这些，给出估计 $P\{N \geqslant k\}$ 的算法如下： 163

步骤 1：生成随机数 $U_1,\cdots,U_m$.

步骤 2：以 $T^{(k)}$ 表示 m 个值 $-\frac{1}{\omega_{n+i}}\log(U_i), i=1,\cdots,m$ 中的第 k 个最大值.

步骤 3：以 $S^{(k)}$ 表示 m 个值 $-\frac{1}{\omega_{n+i}}\log(1-U_i), i=1,\cdots,m$ 中的第 k 个最大值.

步骤 4：模拟得到的估计是

$$\frac{1}{2}\left[\prod_{i=1}^{n}(1-\mathrm{e}^{-\omega_i T^{(k)}}) + \prod_{i=1}^{n}(1-\mathrm{e}^{-\omega_i S^{(k)}})\right] \qquad \square$$

到时刻 t 更新数期望的估计

假设“事件”发生的时间是随机的. 用 T_1 表示第 1 个事件发生的时间，T_2 表示第 1 个与第 2 个事件发生的时间间隔，一般情况，T_n 表示第 $n-1$ 个与第 n 个事件发生的时间间隔，$n \geqslant 1$. 如果令

$$S_n = \sum_{i=1}^{n} T_i$$

则第 1 个事件在时刻 S_1 发生，第 2 个在时刻 S_2 发生. 一般地，第 n 个事件在时刻 S_n 发生（见图 8-2）.

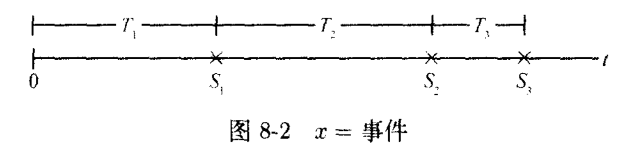


图 8-2　$x=$ 事件

用 $N(t)$ 表示到时刻 t 时发生的事件数，即 $N(t)$ 是时刻 t 前发生的事件数 n 的最大值，或等价地有，

$$N(t)=\max\{n: S_n \leqslant t\}$$

如果相邻事件的时间间隔 $T_1, T_2, \cdots$ 相互独立，且有共同分布 F，则过程 $\{N(t), t \geqslant 0\}$ 被称作*更新过程*.

通过产生中间到达的时间间隔可以容易地模拟更新过程. 假设我们想用模拟方法估计 $\theta=E[N(t)]$，即到某一确定时刻 t 时发生的事件平均数. 我们逐个模拟事件的时间间隔，并同时记录它们的和（代表事件发生的时刻）直到其和超过 t. 也就是说，我们持续产生相邻事件发生的时间间隔，直到 t 之后发生的第一个事件的时刻. 用最初的模拟估计 $N(t)$ 表示到时刻 t 发生的模拟事件数. 我们发现可用作控制变量的一个自然量是产生的 $N(t)+1$ 个时间间隔的序列，也就是说，如果我们用 μ 表示相邻事件发生的时间间隔的均值，则作为随机变量 $T_i-\mu$ 的均值是 0. 于是，

164

$$E\left[\sum_{i=1}^{N(t)+1}\left(T_i-\mu\right)\right]=0$$

因此，我们可用如下形式的估计进行控制：

$$\begin{aligned} N(t)+c\left[\sum_{i=1}^{N(t)+1}\left(T_i-\mu\right)\right] & =N(t)+c\left[\sum_{i=1}^{N(t)+1} T_i-\mu(N(t)+1)\right] \\ & =N(t)+c\left[S_{N(t)+1}-\mu N(t)-\mu\right] \end{aligned}$$

由于 S_n 表示第 n 个事件发生的时间，则 $N(t)+1$ 表示到时刻 t 时已发生事件数加 1，可知 $S_{N(t)+1}$ 表示时间 t 之后第一个事件发生的时间. 因此，如果我们用 $Y(t)$ 记从 t 开始直到下一个事件的时间 [$Y(t)$ 通常被称作在 t 时的剩余寿命]，则

$$S_{N(t)+1}=t+Y(t)$$

于是上面的控制估计可写作

$$N(t)+c[t+Y(t)-\mu N(t)-\mu]$$

其中 c 的最优值由

$$c^* = -\frac{\operatorname{Cov}[N(t), Y(t) - \mu N(t)]}{\operatorname{Var}[Y(t) - \mu N(t)]}$$

给出. 当 t 较大时，可以证明包含 $N(t)$ 的项起主导作用（因为它们的方差随 t 线性增长，考虑到其他项仍有界），于是对较大的 t,

$$c^* \approx -\frac{\operatorname{Cov}[N(t), -\mu N(t)]}{\operatorname{Var}[-\mu N(t)]} = \frac{\mu \operatorname{Var}[N(t)]}{\mu^2 \operatorname{Var}[N(t)]} = \frac{1}{\mu}$$

因此，当 t 较大时，上述形式的最优控制估计接近于

$$N(t) + \frac{1}{\mu}[t + Y(t) - \mu N(t) - \mu] = \frac{Y(t)}{\mu} + \frac{t}{\mu} - 1 \tag{8.7}$$

换句话说，当 t 较大时，用模拟方法确定的临界值是 $Y(t)$，即从时刻 t 开始到下次更新的时间.

165

应用“条件期望”可以对上面的估计做进一步的改进. 我们不用时刻 t 之后第一个事件的实际观测时间，而用最后一个事件发生到时刻 t 的这段时间 $A(t)$（见图 8-3）作条件. 数值 $A(t)$ 通常叫做更新过程在时刻 t 的年龄. [如果我们想象由单个部件构成的系统，部件正常运转的时间是随机的，服从分布 F，而当失效后立即用新部件替换，则得到一个每个事件对应一个部件失效的更新过程. 变量 $A(t)$ 是指在时刻 t 时仍在使用的部件的年龄，这里用年龄意思是指已经使用过的时间数.]

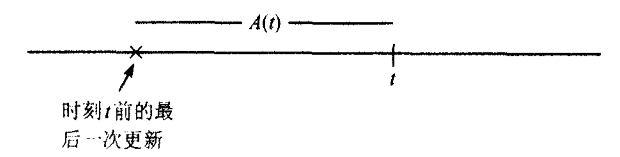


图 8-3 在时刻 t 的年龄

如果过程在时刻 t 的年龄是 x，部件剩余寿命的期望正是在给定部件寿命大于 x 的条件下，部件的使用时间超过 x 的期望数. 也就是，

$$\begin{aligned} E[Y(t)|A(t) = x] &= E[T - x|T > x] \\ &= \int_x^{\infty} (y - x)\frac{f(y)\mathrm{d}y}{1 - F(x)} \equiv \mu[x] \end{aligned}$$

假设上面的 F 是一个连续分布，它的密度函数是 f. 如上定义 $\mu(x)$ 就等于 $E[T - x|T > x]$，由此可得，

$$E[Y(t)|A(t)] = \mu[A(t)]$$

因此，当 t 较大时，得到一个比 (8.7) 式给出的更好的估计：

$$\frac{\mu[A(t)]}{\mu}+\frac{t}{\mu}-1 \tag{8.8}$$

8.4　分层抽样法

假设我们想估计 $\theta = E[X]$，并假设有一个离散随机变量 Y，其可能取值为 $y_1, \cdots, y_k$，并且

(a) 概率 $p_i = P\{Y = y_i\}, i = 1, \cdots, k$ 已知；

166

(b) 对每一个 $i = 1, \cdots, k$，我们可以模拟在 $Y = y_i$ 条件下 X 的值.

现在，如果我们计划通过 n 次模拟估计 $E[X]$，通常使用的方法是产生随机变量 X 的 n 个独立观测值，然后用它们的平均值 $\overline{X}$ 作为 $E[X]$ 的估计. 这个估计的方差是

$$\operatorname{Var}(\overline{X}) = \frac{1}{n}\operatorname{Var}(X)$$

然而，若写成

$$E[X] = \sum_{i=1}^{k} E[X|Y = y_i]p_i$$

则我们可以通过估计 k 个量 $E[X|Y = y_i], i = 1, \cdots, k$ 而得到 $E[X]$ 的估计. 例如，假设我们不是通过产生 n 个独立观测值，而是对每个 $i = 1, \cdots, k$，以每一个事件 $Y = y_i$ 为条件，进行 np_i 次模拟. 如果我们以 $\overline{X}_i$ 作为在 $Y = y_i$ 条件下，产生的 np_i 个观测值的平均值，则我们可以得到 $E[X]$ 的无偏估计

$$\varepsilon = \sum_{i=1}^{k} \overline{X}_i p_i$$

我们称之为 $E[X]$ 的*分层抽样估计*.

因为 $\overline{X}_i$ 是 np_i 个独立随机变量的平均值，这些变量的分布等同于在给定 $Y = y_i$ 下 X 的条件分布，于是

$$\operatorname{Var}(\overline{X}_i) = \frac{\operatorname{Var}(X|Y = y_i)}{np_i}$$

因此，根据前述方法以及 $\overline{X}_i, i = 1, \cdots, k$ 之间的独立性，有

$$\operatorname{Var}(\varepsilon) = \sum_{i=1}^{k} p_i^2 \operatorname{Var}(\overline{X}_i)$$

$$= \frac{1}{n}\sum_{i=1}^{k} p_i \text{Var}(X|Y=y_i)$$
$$= \frac{1}{n}E[\text{Var}(X|Y)]$$

由于 $\text{Var}(\overline{X}) = \frac{1}{n}\text{Var}(X)$，$\text{Var}(\varepsilon) = \frac{1}{n}E[\text{Var}(X|Y)]$，通过条件方差公式，得到

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

167

我们看到应用分层抽样估计 ε 比通常的模拟估计的方差减小了

$$\text{Var}(\overline{X}) - \text{Var}(\varepsilon) = \frac{1}{n}\text{Var}(E[X|Y])$$

也就是说，当 Y 值明显地影响 X 的条件期望时，每一次运行节省的方差 $\text{Var}(E[X|Y])$ 是很有价值的.

注 如以 S_i^2 表示在给定 $Y = y_i, i = 1, \cdots, k$ 的条件下，np_i 次模拟运行的样本方差，则我们可以用此方差估计分层抽样估计的方差. 事实上，S_i^2 是 $\text{Var}(X|Y=y_i)$ 的无偏估计，由此得到 $\text{Var}(\varepsilon)$ 的一个无偏估计 $\frac{1}{n}\sum_{i=1}^{k} p_i S_i^2$.

例 8q 在一个有无限多个服务员的排队系统中，晴天顾客到达服从参数为每小时 12 个的泊松过程，而在其他天气条件下，顾客到达服从参数为每小时 4 个的泊松过程. 在所有的时间内，服务时间服从参数是每小时 1 个的指数分布. 每天系统工作 10 小时后关闭，所有那些正在系统中而没有完成服务的顾客被迫离去. 假设某天是晴天的概率为 0.5 并且与其他天独立，则我们想用模拟方法估计每天没有完成服务的顾客数的均值 θ.

用 X 表示在随机选取的一天中服务没有完成的顾客数；用 $y = 0$ 表示天气一般，用 $y = 1$ 表示天气好. 可以证明在给定 $Y = 0$ 和 $Y = 1$ 时，X 的条件分布都服从泊松分布，均值分别是

$$E[X|Y=0] = 4(1-\mathrm{e}^{-10}), \qquad E[X|Y=1] = 12(1-\mathrm{e}^{-10})$$

由于泊松随机变量的方差等于它的均值，故

$$\text{Var}(X|Y=0) = E[X|Y=0] \approx 4$$
$$\text{Var}(X|Y=1) = E[X|Y=1] \approx 12$$

因此，

$$E[\text{Var}(X|Y)] \approx \frac{1}{2}(4+12) = 8$$

和

$$\text{Var}(E[X|Y]) = E[(E[X|Y])^2] - (E[X])^2 \approx \frac{1}{2}[(4^2 + (12)^2] - 8^2 = 16$$

168

因此

$$\mathrm{Var}(X) \approx 8 + 16 = 24$$

这个估计的方差大约是以好天气实际占一半，一般天气实际占一半进行分层抽样所得估计的方差 $E[\mathrm{Var}(X|Y)]$ 的 3 倍. □

如再假设概率分布函数 $p_i = P\{Y = y_i\}, i = 1, \cdots, k$ 已知，则我们可以在条件 $Y = i$ 下，模拟 X，并且计划做 n 次模拟运行. 虽然在条件 $Y = y_i$，$i = 1, \cdots, k$ 下所做 n 次模拟中的 n_i 次比产生 X 的 n 个独立重复观测值要好，但这并不是完成条件模拟所需要的最优次数. 假设我们计划在条件 $Y = y_i$ 下做 n_i 次模拟，其中 $n = \sum_{i=1}^{k} n_i$. 若用 $\overline{X}_i$ 表示在条件 $Y = y_i$ 下，n_i 次模拟的平均值，则分层抽样估计是

$$\hat{\theta} = \sum_{i=1}^{k} p_i \overline{X}_i$$

该估计的方差为

$$\mathrm{Var}(\hat{\theta}) = \sum_{i=1}^{k} p_i^2 \mathrm{Var}(X|Y = i)/n_i$$

尽管

$$\mathrm{Var}(X|Y = i),\ i = 1, \cdots, k$$

的值可能开始并不知道，但我们可以通过少量的模拟研究估计它们，比如用估计量 S_i^2. 我们可以通过求解如下的最优化问题选择 n_i：

$$\text{在 } \sum_{i=1}^{k} n_i = n \text{ 条件下，适当地选取 } n_1, \cdots, n_k, \text{使 } \sum_{i=1}^{k} p_i^2 S_i^2/n_i \text{ 达到最小.}$$

利用拉格朗日乘子法，容易证明上述最优问题的解为

$$n_i = n \frac{p_i s_i}{\sum_{j=1}^{k} p_j s_j},\ i = 1, \cdots, k$$

一旦 n_i 被选定并且模拟完成，我们将用 $\sum_{i=1}^{k} p_i \bar{X}_i$ 估计 $E[X]$，用 $\sum_{i=1}^{k} p_i^2 S_i^2/n_i$ 估计这个估计的方差，其中 S_i^2 是在 $Y = y_i, i = 1, \cdots, k$ 条件下，n_i 次运行得到的样本方差.

169

下面我们再举例说明分层抽样法的优点，假设想通过 n 次模拟估计

$$\theta = E[h(U)] = \int_0^1 h(x)\mathrm{d}x$$

若令

$$S = j,\ \text{如果} \frac{j-1}{n} \leqslant U < \frac{j}{n},\ j = 1, \cdots, n$$

则

$$\theta = \frac{1}{n}\sum_{j=1}^{n} E[h(U)|S=j\,]$$
$$= \frac{1}{n}\sum_{j=1}^{n} E[h(U_{(j)})]$$

其中 $U_{(j)}$ 是 $((j-1)/n, j/n)$ 上均匀分布的随机变量. 故由上述可知，我们不使用 $\sum_{j=1}^{n} h(U_j)/n$ 估计 θ，而是通过产生 $U_1,\cdots,U_n$ 给出 θ 的一个更好的估计

$$\widehat{\theta} = \frac{1}{n}\sum_{j=1}^{n} h\left(\frac{U_j+j-1}{n}\right)$$

例 8r 在例 8k 中，我们证明了

$$\frac{\pi}{4} = E\left[\sqrt{(1-U^2)}\right]$$

因此，我们可以通过产生 $U_1,\cdots,U_n$ 估计 π，并用

$$\text{est} = \frac{4}{n}\sum_{j=1}^{n}\sqrt{1-[(U_j+j-1)/n]^2}$$

作为 π 的估计. 实际上，我们可以使用对偶变量得到一个进一步改进的估计

$$\widehat{\pi} = \frac{2}{n}\sum_{j=1}^{n}\left(\sqrt{1-[(U_j+j-1)/n]^2}+\sqrt{1-[(j-U_j)/n]^2}\right)$$

170

关于估计量 $\widehat{\pi}$ 的模拟结果如下：

n	$\widehat{\pi}$
5	3.161211
10	3.148751
100	3.141734
500	3.141615
1000	3.141601
5000	3.141593

当 $n=5000$ 时，估计 $\widehat{\pi}$ 校正到小数点后六位. □

注 1 假设我们想通过对 Y 的值进行分层抽样模拟估计 $E[X]$，其中 Y 是分布函数为 G 的连续随机变量. 为完成分层抽样，我们首先产生 Y 的值，然后以 Y

的这个值为条件，模拟 X. 我们使用逆变换方法产生 Y；也就是说，通过产生随机数 U，并且令 $Y=G^{-1}(U)$ 得到 Y 的值. 如果我们计划做 n 次模拟，则我们不是用这 n 个独立的随机数产生 Y 的值，而是令第 i 个随机数取在区间 $(\frac{i-1}{n},\frac{i}{n})$ 上服从均匀分布的随机变量的值进行分层抽样. 以这种方式，通过产生随机数 U_i，并且令 $Y_i=G^{-1}(\frac{U_i+i-1}{n})$ 得到第 i 次运行产生的 Y 值 Y_i. 通过在给定 Y 等于 Y_i 的观测值的条件下，模拟 X，得到第 i 次模拟运行对应的 X 的值 X_i. 由于随机变量 X_i 是 $E[X|G^{-1}(\frac{i-1}{n})<Y\leqslant G^{-1}(\frac{i}{n})]$ 的无偏估计，则知 $\frac{1}{n}\sum_{i=1}^{n}X_i$ 是

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}E\left[X\middle|G^{-1}\left(\frac{i-1}{n}\right)<Y\leqslant G^{-1}\left(\frac{i}{n}\right)\right]\\
&=\sum_{i=1}^{n}E\left[X\middle|\frac{i-1}{n}<G(Y)\leqslant\frac{i}{n}\right]\frac{1}{n}\\
&=\sum_{i=1}^{n}E\left[X\middle|\frac{i-1}{n}<G(Y)\leqslant\frac{i}{n}\right]P\left\{\frac{i-1}{n}<G(Y)\leqslant\frac{i}{n}\right\}\\
&=E[X]
\end{aligned}
$$

的一个无偏估计，其中倒数第二个等式应用了 $G(Y)$ 是 $(0,1)$ 上的均匀分布.

171 **注 2** 假设没有使用分层抽样，我们已经模拟出 n 个 X 的独立重复值，并且在 $n_i(\sum_{i=1}^{k}n_i=n)$ 次模拟中，得到 Y 的值是 y_i. 如果我们用 $\overline{X}_i$ 表示 n_i 次运行的平均值，对应 Y 的值是 y_i，则所有 n 次运行的平均值 $\overline{X}$ 可以表示为

$$
\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}\sum_{i=1}^{k}n_i\overline{X}_i=\sum_{i=1}^{k}\frac{n_i}{n}\overline{X}_i
$$

当改写成这种形式，很明显用 $\overline{X}$ 估计 $E[X]$ 等价于对每个 $i=1,\cdots,k$ 用 $\overline{X}_i$ 估计 $E[X|Y=i]$ 和用 n_i/n 估计 p_i. 但是因为 p_i 是已知的，所以不需估计，可知 $\sum_{i=1}^{k}p_i\overline{X}_i$ 作为 $E[X]$ 的估计比 $\overline{X}$ 要好. 换句话说，如同在 y_i, $i=1,\cdots,k$ 的条件下，已经做了 n_i 次模拟运行一样，我们应该做分层抽样. 这种事后做分层抽样的方法叫做*后验分层抽样*(poststratification). □

再次假设我们感兴趣的是估计 $\theta=E[X]$，其中 X 依赖于随机变量 S，而 S 取值 $1,\cdots,k$ 的概率分别是 p_i. 则

$$
E[X]=p_1E[X|S=1]+p_2E[X|S=2]+\cdots+p_kE[X|S=k]
$$

如果所有的 $E[X|S=i]$ 都是已知的（即：如果 $E[X|S]$ 已知），但是 p_i 是未知的，则可以通过产生 S 的值，用条件期望 $E[X|S]$ 估计 θ. 另一方面，如果 p_i 是已知的，可以得到在 S 的值给定时，X 的条件分布，通过模拟得到 $E[X|S=i]$ 的估计

$\widehat{E}[X|S=i]$，然后用分层抽样估计 $\sum_{i=1}^{k}\widehat{E}[X|S=i]$ 估计 $E[X]$. 当部分 p_i 与部分 $E[X|S=i]$ 已知时，可以将这些方法联合使用. □

例 8s 在视频纸牌游戏中，玩家将一枚硬币塞入机器中，机器随机发给玩家一手牌共五张，玩家可以放弃其中几张，从剩余的 47 张中抓新牌替换. 庄家根据玩家手中牌的组合返还给玩家一定数量的硬币. 下面是一典型的收益表： 172

牌	收 益
同花大顺	800
同花顺	50
四张相同牌	25
三张相同的牌和另一个对	8
同花	5
顺	4
三张相同的牌	3
两个对	2
J 以上的对	1
什么也没有	0

一手牌如果它确实属于上面的某个类型并且它不是更高的类型，则就认为是上面的这个类型. 就是说，例如，同花意思是五张牌同花色，但不全连在一起.

考虑如下的策略：如果最初的牌含有顺或更好的组合，就不再补牌 (也就是说，玩家坚持不变)，并且通常保留对或三张相同的牌. 在这种策略下，用 X 表示玩家在单手牌中赢利，并假设我们感兴趣的是估计 $\theta=E[X]$. 我们以最初发到玩家手中牌的类型为条件开始，而不直接用 X 做估计. 用 R 代表"同花大顺"，S 代表"同花顺"，4 代表"四张相同的牌"，3 代表"三张相同的牌"，2 代表"两个对"，1 代表"J 以上的对"，0 代表"小对"，"other" 代表"其他未提到的任何牌型"，F 表示"三张相同和另一对牌"，则可得

$$
\begin{aligned}
E[X]=&E[X|R]P\{R\}+E[X|S]P\{S\}+E[X|4]P\{4\}+E[X|F]P\{F\}\\
&+E[X|\text{同花}]P\{\text{同花}\}+E[X|\text{顺}]P\{\text{顺}\}+E[X|3]P\{3\}\\
&+E[X|2]P\{2\}+E[X|1]P\{1\}+E[X|0]P\{0\}+E[X|\text{other}]P\{\text{other}\}
\end{aligned}
$$

记 $C=\binom{52}{5}^{-1}$，可得

$$
\begin{aligned}
P\{R\}&=4C=1.539\times10^{-6}\\
P\{S\}&=4\times9\times C=1.3851\times10^{-4}\\
P\{4\}&=13\times48\times C=2.400\,96\times10^{-4}
\end{aligned}
$$

$$P\{F\}=13\times 12\binom{4}{3}\binom{4}{2}C=1.440\ 576\times 10^{-3}$$

173

$$P\{\text{同花}\}=4\left(\binom{13}{5}-10\right)C=1.965\ 402\times 10^{-3}$$

$$P\{\text{顺}\}=10(4^5-4)C=3.924\ 647\times 10^{-3}$$

$$P\{3\}=13\binom{12}{2}4^3C=2.112\ 845\ 1\times 10^{-2}$$

$$P\{2\}=\binom{13}{2}44\binom{4}{2}\binom{4}{2}C=4.753\ 901\ 6\times 10^{-2}$$

$$P\{1\}=4\binom{4}{2}\binom{12}{3}4^3C=0.130\ 021\ 239$$

$$P\{0\}=9\binom{4}{2}\binom{12}{3}4^3C=0.292\ 547\ 788$$

$$\begin{aligned}P\{\text{other}\}=&1-P\{R\}-P\{S\}-P\{F\}-P\{\text{同花}\}\\&-P\{\text{顺}\}-\sum_{i=0}^{4}P\{i\}=0.501\ 052\ 7\end{aligned}$$

于是，得到

$$E[X]=0.051\ 290\ 3+\sum_{i=0}^{3}E[X|i]P\{i\}+E[X|\text{other}]0.501\ 052\ 7$$

通过如下分析可以计算出 $E[X|3]$，首先注意到两张新牌来自剩余的 47 张中，它包括 1 个单张，2 个 3 张一样大的，10 个 4 张一样大的. 如此，用 H 表示最后手中的牌型，则有

$$P\{H=4|3\text{ 张点数相同}\}=\frac{46}{\binom{47}{2}}=0.042\ 553\ 191$$

$$P\{H=F|3\text{ 张点数相同}\}=\frac{2\times 3+10\times 6}{\binom{47}{2}}=0.061\ 054\ 579$$

$$P\{H=3|3\text{ 张点数相同}\}=1-0.042\ 553\ 191-0.061\ 054\ 579=0.896\ 392\ 23$$

因此,

$$E[X|3]=25(0.042\ 553\ 191)+8(0.061\ 054\ 579)+3(0.896\ 392\ 23)$$
$$=4.241\ 443\ 097$$

当 $i=0,1,2$ 时，可类似地推导出 $E[X|i]$（具体推导留作习题）.

在模拟运行中，应该按如下的方式产生一手牌. 如果它包含至少一对或更高的牌，则就将它放弃，并使过程再次开始. 当发到的牌不包含对（或更好的牌），应用可使用的任何策略垫牌并接受新牌. 如果 X_o 是这一手的收益，则 X_o 是 $E[X|\text{other}]$ 的估计，并且基于这次模拟，θ 的估计是 174

$$\widehat{\theta}=0.051\ 290\ 3+0.021\ 128\ 451(4.241\ 443\ 097)+0.047\ 539\ 016E[X|2]$$
$$+0.130\ 021\ 239E[X|1]+0.292\ 547\ 788E[X|0]+0.501\ 052\ 7X_o$$

注意到这个估计的方差是

$$\text{Var}(\widehat{\theta})=(0.501\ 052\ 7)^2\text{Var}(X_o)$$ □

注 1 已经假设采取的策略是：若是一手合适的牌则通常不换牌，并且通常情况下任何对都保留. 然而，当报酬给定时，这并不是最优策略. 例如，如果一手牌是 2，10，J，Q，K，并且都是黑桃，采取的策略不是保留这个同花，更好的策略是垫去 2 再另抓一张（这是为什么？）. 还有如果一手牌中含有黑桃 10, J, Q, K，另一张是红桃 10，更好的策略是垫去红桃 10，再另抓一张而不是保留对 10.

注 2 可以通过将“其他”类分成包含有 4 张同一花色的“其他”类与不包含有 4 张同一花色的“其他”类的方式进一步应用分层抽样法. 在分析上计算一手牌是无对而有 4 张同一花色的这种“其他”类的概率并不困难. 可以用模拟估计这两种“其他”类收益的条件期望. □

8.5 分层抽样法的应用

在下面的小节中，我们介绍在分析以泊松到达的系统，多变量的单调函数和复合随机向量时，如何使用分层抽样方法.

在 8.5.1 节我们考虑到达依据泊松过程的模型，并给出一个有效地估计随机变量期望值的方法，其中该随机变量的期望仅通过在某一给定时间到达者人数总和依赖该到达过程. 在 8.5.2 节我们指出如何利用分层抽样方法有效地估计随机数的非减函数的期望. 在 8.5.3 节我们定义复合随机向量的概念，并且给出如何有效地估计这个向量的函数的期望. 175

到达服从泊松过程的系统分析

考虑一个依据泊松过程到达的系统，假设我们感兴趣的是通过模拟计算 $E[D]$，其中 D 的值仅通过时间 t 之前的到达者依赖该过程. 例如，D 可能是在一个并联多服务排队系统中到时间 t 为止的所有到达者延迟的时间总和. 我们建议用下面方法模拟估计 $E[D]$. 首先，$N(t)$ 等于到时间 t 为止到达者个数，注意到对任意指定的整数 m

$$E[D]=\sum_{j=0}^{m}E[D|N(t)=j]\mathrm{e}^{-\lambda t}(\lambda t)^j/j!+E[D|N(t)>m]\times\left(1-\sum_{j=0}^{m}\mathrm{e}^{-\lambda t}(\lambda t)^j/j!\right) \quad (8.9)$$

我们假设 $E[D|N(t)=0]$ 容易计算，并且 D 的值由已知到达时间和每个到达者接受服务的时间共同确定.

用我们提供的模拟方法，每运行一次产生一个独立的 $E[D]$ 的估计. 进一步来说，每次运行由 $m+1$ 步组成，对 $j=1,\cdots,m$，第 j 步产生 $E[D|N(t)=j]$ 的一个无偏估计，第 $m+1$ 步产生 $E[D|N(t)>m]$ 的一个无偏估计. 接下来的每一个步骤使用其前一步提供的数据和另外一些需要的数据，这些数据分别是在第 $2,\cdots,m$ 中的到达时间和接受服务的时间. 为了跟踪当前的到达时间，每一步将有一个对应的集合 S，其元素表示到达时间且按增序排列. 为了从上一步到下一步，我们应用如下事实：以时刻 t 到达的总数 j 为条件，j 个到达时间的集合的分布与 $(0,t)$ 上 j 个独立的均匀分布的随机变量的分布一样. 因此以时刻 t 的 $j+1$ 个事件为条件，到达时间的集合的分布等价于以时刻 t 的 j 个事件和一个新的独立的 $(0,t)$ 上均匀分布的随机变量构成事件的分布一样.

模拟方法如下：

步骤 1：令 $N=1$. 生成随机数 U_1，并令 $S=\{tU_1\}$.

步骤 2：假设 $N(t)=1$，到达发生在时间 tU_1，生成这个到达者的服务时间，并计算 D 的值，称这个值是 D_1.

步骤 3：令 $N=N+1$.

步骤 4：生成随机数 U_N，并将 tU_N 放入集合 S 的适当位置，使其元素按增序排列.

176

步骤 5：假设 $N(t)=N$，S 表明 N 个到达的时间；生成在 tU_N 到达者的服务时间，并利用以前产生的其他到达者的服务时间，计算 D 的值，称这个值是 D_N.

步骤 6：如果 $N<m$ 返回步骤 3. 如果 $N=m$，以它的值超过 m 为条件，用逆变换方法生成 $N(t)$ 的值. 如果生成的 $N(t)$ 的值是 $m+k$. 再生成 k 个另外的随机数，每个都乘以 t，并将这 k 个数添加到 S 中. 生成这 k 个到达者的服务时间，并结合以前产生的其他到达者的服务时间，计算 D 的值，称这个值是 $D_{>m}$.

记 $D_0 = E[D|N(t)=0]$，这次模拟的估计是

$$\varepsilon = \sum_{j=0}^{m} D_j e^{-\lambda t}(\lambda t)^j/j\,! + D_{>m}\left(1 - \sum_{j=0}^{m} e^{-\lambda t}(\lambda t)^j/j\,!\right) \tag{8.10}$$

因为在给定 $N(t)=j$ 条件下，无序到达时间集合的分布与 $(0,\ t)$ 上 j 个独立的均匀分布的随机变量集合的分布一样，于是

$$E[D_j] = E[D|N(t)=j], \quad E[D_{>m}] = E[D|N(t)>m]$$

由此知道 ε 是 $E[D]$ 的一个无偏估计. 进行多次模拟运行并计算估计结果的平均值得到最后的模拟估计.

注 1 应注意到因为重复使用相同的数据导致的正相关，我们的估计 $\sum_{j=0}^{m} D_j e^{-\lambda t}(\lambda t)^j/j\,! + D_{>m}(1-\sum_{j=0}^{m} e^{-\lambda t}(\lambda t)^j/j\,!)$ 的方差比 D_j 是独立的估计的方差大. 然而，其模拟速度的加快能弥补方差的增大.

注 2 计算 D_{j+1} 时，可以利用计算 D_j 时的结果. 例如，假设当 $N(t)=j$ 时，第 i 个到达者延迟的时间是 $D_{i,j}$. 如果新的到达者的时间 tU_{j+1} 是新集合 S 中的第 k 个最小值，则对 $i<k$，有 $D_{i,j+1}=D_{i,j}$.

注 3 其他的缩减方差的技术可以和我们这里的方法联合使用. 例如，我们可以用服务时间的线性组合作为控制变量改进这个估计. □ 177

接下来是确定 m 的合适的取值. 合理的方法是选取 m 使

$$E[D|N(t)>m]P\{N(t)>m\} = E[D|N(t)>m]\left(1 - \sum_{j=0}^{m} e^{-\lambda t}(\lambda t)^j/j\,!\right)$$

充分地小. 因为 $\mathrm{Var}(N(t)) = \lambda t$，合理的选择应该有如下的形式

$$m = \lambda t + k\sqrt{\lambda t},$$

其中 k 是某一正数.

为确定 k 的合适的取值，我们可以先给出 $E[D|N(t)>m]$ 的上界，然后利用这个界确定 k(和 m) 的合适的取值. 例如，D 可以是一个平均服务时间是 1 的单服务系统中时刻 t 所有到达者延迟时间的总和. 因为当所有到达者都同时到达时，这个和达到最大值，我们看到

$$E[D|N(t)] \leqslant \sum_{i=1}^{N(t)-1} i$$

因为当 m 至少大于 $E[N(t)]$ 的 5 个标准差时，在给定 $N(t)$ 大于 m 的条件下，其条件分布主要集中在 $m+1$ 附近，由此我们可以合理假设，当 $k \geqslant 5$ 时，

$$E[D|N(t)>m] \leqslant (m+1)^2/2$$

对于标准正态随机变量 Z，有 [1]

$$P(Z>x) \leqslant (1-1/x^2+3/x^4)\frac{\mathrm{e}^{-x^2}/2}{x\sqrt{2\pi}}, \qquad x>0$$

对泊松分布用正态近似，我们看到，当 $k \geqslant 5$ 且 $m=\lambda t+k\sqrt{\lambda t}$ 时，可以合理假设

$$E[D|N(t)>m]P\{N(t)>m\} \leqslant (m+1)^2\frac{\mathrm{e}^{-k^2}/2}{2k\sqrt{2\pi}}$$

例如，取 $\lambda t=10^3$ 和 $k=6$，上面的上界大约是 0.0008.

178 我们通过证明估计 ε 的方差小于原始模拟估计 D 的方差来结束这一小节.

定理

$$\mathrm{Var}(\varepsilon) \leqslant \mathrm{Var}(D)$$

证明 我们通过证明 ε 可以表示为在给定某随机向量的条件下的 D 的条件期望来证明定理的结论. 为证明这点，我们利用如下的方法模拟 D：

步骤 1. 生成 N' 的值，其中 N' 是与 $N(t)$ 在 $N(t)>m$ 条件下的条件分布同分布的随机变量. 也就是说，

$$P\{N'=k\}=\frac{(\lambda t)^k/k\,!}{\sum_{k=m+1}^{\infty}(\lambda t)^k/k\,!}, \quad k>m$$

步骤 2. 生成 $A_1,\cdots,A_{N'}$ 的值，它们是相互独立的随机变量，都服从 $(0,t)$ 上均匀分布.

步骤 3. 生成 $S_1,\cdots,S_{N'}$ 的值，它们是相互独立的服务时间随机变量.

步骤 4. 生成 $N(t)$ 的值，其中 $N(t)$ 是均值为 λt 的泊松随机变量.

步骤 5. 如果 $N(t)=j \leqslant m$，用到达时间 $A_1,\cdots,A_j$ 和服务时间 $S_1,\cdots,S_j$ 计算 $D=D_j$ 的值.

步骤 6. 如果 $N(t)>m$，用到达时间 $A_1,\cdots,A_{N'}$ 和服务时间 $S_1,\cdots,S_{N'}$ 计算 $D=D_{>m}$ 的值.

1. 见 Sec. 3.3 of Ross, *Probability Models for Computer Science*, Academic Press, 2002.

注意到,

$$
\begin{aligned}
&E[D|N', A_1, \cdots, A_{N'},\ \ S_1, \cdots, S_{N'}\,] \\
=&\sum_j E[D|N',\ A_1, \cdots, A_{N'},\ \ S_1, \cdots, S_{N'}, N(t)=j] \\
&\times P\{N(t)=j|N',\ A_1, \cdots, A_{N'},\ \ S_1, \cdots, S_{N'}\,] \\
=&\sum_j E[D|N',\ A_1, \cdots, A_{N'},\ \ S_1, \cdots, S_{N'}, N(t)=j]P\{N(t)=j\} \\
=&\sum_{j=0}^{m} D_j P\{N(t)=j\} + \sum_{j>m} D_{>m} P\{N(t)=j\} \\
=&\,\varepsilon
\end{aligned}
$$

我们看到 ε 是给定某些数据下 D 的条件期望. 因此, 根据条件方差公式结论成立. □ [179]

计算单调函数的多重积分

假设我们想要用模拟方法估计 n 重积分

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 g(x_1, x_2, \cdots, x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$$

假设 $U_1, \cdots, U_n$ 是服从 $(0,1)$ 上均匀分布的相互独立的随机变量, 上式可表示为

$$\theta = E[g(U_1, \cdots, U_n)]$$

假设 g 是其自变量的非减函数. 也就是说, 对每个 $i=1,\cdots,n$ 及固定的 $x_1,\cdots,x_{i-1}$, $x_{i+1},\cdots,x_n$, 函数 $g(x_1,\cdots,x_i,\cdots,x_n)$ 是 x_i 的增函数. 如果我们令 $Y=\prod_{i=1}^n U_i$, 则因为 Y 和 $g(U_1,\cdots,U_n)$ 都是 U_i 的增函数, 看起来 $E[\mathrm{Var}(g(U_1,\cdots,U_n)|Y)]$ 可能经常会相当地小. 因此, 我们应该考虑通过对 $\prod_{i=1}^n U_i$ 进行分层估计 θ. 为了实现这个目标, 我们需要先说明:

(a) 在 $\prod_{i=1}^n U_i$ 的值确定的条件下, 如何生成 $U_1,\cdots,U_n$ 的值;

(b) 在分层模式中, 如何生成 $\prod_{i=1}^n U_i$ 的值.

为了解决上述两个问题, 我们将 U_i 与泊松过程建立联系. 注意到 $-\ln(U)$ 是参数为 1 的指数型随机变量, 且可以把 $-\ln(U_i)$ 解释为参数 1 的泊松过程第 $i-1$ 与第 i 个事件之间的间隔时间. 于是, 泊松过程第 j 个事件将在时间 T_j 发生, 其中

$$T_j = \sum_{i=1}^{j} -\ln(U_i) = -\ln(U_1\cdots U_j)$$

因为 n 个参数为 1 的独立的指数型随机变量的和是一个服从 $\Gamma(n,1)$ 分布的随机变量，于是，通过产生 (在分层模式中讨论) 服从 $\Gamma(n,1)$ 分布的随机变量，就可以得到 $T_n = -\ln(U_1\cdots U_n)$ 的值. 这样生成了 $\prod_{i=1}^n U_i$ 的值. 也就是说，

$$\prod_{i=1}^{n} U_i = \mathrm{e}^{-T_n}$$

180

在给定 n 个随机变量 $U_1,\cdots,U_n$ 的乘积条件下，为生成其每一个值，我们应用泊松过程的如下结论：在已知其第 n 个事件出现在时刻 t 条件下，其前 $n-1$ 个事件即为排序后的相互独立的 $(0,t)$ 上均匀分布的随机变量. 于是，一旦生成 T_n 的值，单个 U_i 的值可按如下的方式得到：生成 $n-1$ 个随机数 $V_1,\cdots,V_{n-1}$，排序后得到 $V_{(1)}<V_{(2)}<\cdots<V_{(n-1)}$. 由于 $T_nV_{(j)}$ 表示第 j 个事件发生的时间，则有

$$\begin{aligned} T_nV_{(j)} &= -\ln(U_1\cdots U_j) \\ &= -\ln(U_1\cdots U_{j-1}) - \ln(U_j) \\ &= T_nV_{(j-1)} - \ln(U_j) \end{aligned}$$

因此，由 $V_{(0)}=0$, $V_{(n)}=1$，得

$$U_j = \mathrm{e}^{-T_n[V_{(j)}-V_{(j-1)}]},\ j=1,\cdots,n \tag{8.11}$$

于是，我们看到如何在给定 $\prod_{i=1}^n U_i$ 的值的条件下，生成随机变量 $U_1,\cdots,U_n$. 为完成分层抽样，我们应用 $T_n = -\ln(\prod_{i=1}^n U_i)$ 是一个服从 $\Gamma(n,1)$ 分布的随机变量这样一个事实. 令 G_n 表示服从 $\Gamma(n,1)$ 分布的随机变量的分布函数. 如果我们计划做 m 次模拟运行，则第 k 次运行应该产生一个随机数 U, T_n 取值应该等于 $G_n^{-1}(\frac{U+k-1}{m})$. 对于 T_n 这个值，我们应用上述的公式模拟生成 $U_1,\cdots,U_n$ 的值，并计算 $g(U_1,\cdots,U_n)$. [也就是说，我们产生 $n-1$ 个随机数并对它们排序，得到 $V_{(1)}<V_{(2)}<\cdots<V_{(n-1)}$，利用 (8.11) 式得到 U_j]. 这 m 次运行得到的 g 的值的平均就是 $E[g(U_1,\cdots,U_n)]$ 的分层抽样估计.

注 1　参数是 $(n,1)$ 的伽玛随机变量与 $\frac{1}{2}\chi_2^2$ 的分布相同，其中 χ_{2n}^2 是自由度为 $2n$ 的卡方随机变量. 因此，

$$G_n^{-1}(x) = \frac{1}{2}F_{\chi_{2n}^2}^{-1}(x)$$

其中 $F_{\chi_{2n}^2}^{-1}(x)$ 是自由度为 $2n$ 的卡方随机变量的分布函数的逆. 在文献中，卡方分布的逆函数的近似值经常会用到.

注 2　即使考虑的函数是其中某些自变量的单调增函数和其余自变量的单调

181

减函数，我们做适当的修正，也可以应用上述的分层抽样方法. 例如，假设我们想

求 $E[h(U_1,\cdots,U_n)]$ 的值，其中 h 是其第一个自变量的单调减函数，其他自变量的单调增函数. 利用 $1-U_1$ 也是 $(0,1)$ 上均匀分布的随机变量，可以得出

$$E[h(U_1,U_2\cdots,U_n)] = E[h(1-U_1,U_2\cdots,U_n)] = E[g(U_1,U_2\cdots,U_n)]$$

其中 $g(x_1,x_2,\cdots,x_n) \equiv h(1-x_1,x_2,\cdots,x_n)$ 是其每一个自变量的单调增函数. □

复合随机向量

以 N 表示非负整数值随机变量，其概率分布函数是

$$p(n) = P\{N=n\}$$

并假设 N 与另一组来自分布函数 F 的独立随机变量序列 $X_1,X_2,\cdots$ 相互独立. 则随机向量 $(X_1,X_2,\cdots,X_N)$ 叫做复合随机向量.(当 $N=0$ 时，称这个复合随机向量为零向量.)

对于函数族 $g_n(x_1,x_2,\cdots,x_n)$, $n\geqslant 0$, 其中 $g_0\equiv 0$, 对指定的随机向量 $(X_1,X_2,\cdots,X_N)$, 假设我们感兴趣的是模拟估计 $E[g_N(X_1,X_2,\cdots,X_N)]$. 以下是一些我们感兴趣的函数族.

- 如果

$$g_n(x_1,x_2,\cdots,x_n) = \begin{cases} 1 & \text{如果} \sum_{i=1}^n x_i > a \\ 0 & \text{其他} \end{cases}$$

  则 $E[g_N(X_1,X_2,\cdots,X_N)]$ 是复合随机向量超过 a 的概率.

- 上面例子的一般情况是：对 $0<\alpha<1$,

$$g_n(x_1,x_2,\cdots,x_n) = \begin{cases} 1 & \text{如果} \sum_{i=1}^n \alpha^i x_i > a \\ 0 & \text{其他} \end{cases}$$

  则 $E[g_N(X_1,X_2,\cdots,X_N)]$ 是复合随机向量的折扣和超过 a 的概率.

- 上述两个例子是如下一般情况的特例：对给定的序列 $a_i, i\geqslant 1$,

$$g_n(x_1,x_2,\cdots,x_n) = \begin{cases} 1 & \text{如果} \sum_{i=1}^n a_i x_i > a \\ 0 & \text{其他} \end{cases}$$

182

- 有时，我们感兴趣的是随机向量的前 k 个最大值的加权和的函数，这促使我们考虑函数

$$g_n(x_1,x_2,\cdots,x_n) = g\left(\sum_{i=1}^{\min(k,n)} a_i x_{(i:n)}\right)$$

其中 $X_{(i:n)}$ 是 $x_1, x_2, \cdots, x_n$ 中的第 i 个最大值，其中 g 是确定的函数，且 $g(0)=0$.

为模拟估计 $\theta \equiv E[g_N(X_1, X_2, \cdots, X_N)]$，我们选取值 m 使 $P\{N>m\}$ 很小，且假设在 N 超过 m 的条件下，我们能够模拟 N. 如记 $p_n = P\{N=n\}$，由于 N 仅取值 $0, 1, \cdots, m$ 中的一个或超过 m，且这些取值是相互独立的，于是

$$\begin{aligned}
\theta &= \sum_{n=0}^{m} E[g_N(X_1, X_2, \cdots, X_N)|N=n]p_n + E[g_N(X_1, X_2, \cdots, X_N)|N>m]P\{N>m\} \\
&= \sum_{n=0}^{m} E[g_n(X_1, X_2, \cdots, X_n)|N=n]p_n + E[g_N(X_1, X_2, \cdots, X_N)|N>m]P\{N>m\} \\
&= \sum_{n=0}^{m} E[g_n(X_1, X_2, \cdots, X_n)]p_n + E[g_N(X_1, X_2, \cdots, X_N)|N>m]\left[1-\sum_{n=0}^{m} p_n\right]
\end{aligned}$$

其中最后一个等式应用了 N 与 $X_1, \cdots, X_n$ 之间的独立性.

为了有效地模拟估计 $E[g_N(X_1, X_2, \cdots, X_N)]$，首先在 N 超过 m 的条件下，模拟 N 的值. 假设产生的值是 m'. 接着，产生 m' 个独立的随机变量 $X_1, \cdots, X_{m'}$，它们有共同的分布函数 F. 这样完成一次模拟运行，该次模拟对应的估计是

$$\varepsilon = \sum_{n=1}^{m} g_n(X_1, X_2, \cdots, X_n)p_n + g_{m'}(X_1, X_2, \cdots, X_{m'})\left[1-\sum_{n=0}^{m} p_n\right]$$

注 1　如果计算函数 g_n 的值相当容易，我们建议也可以用 $X_1, \cdots, X_{m'}$ 的逆序得到第二个估计，并对这两个估计取平均. 也就是说，用如下的估计

183

$$\varepsilon^* = \frac{1}{2}\left(\varepsilon + \sum_{n=1}^{m} g_n(X_{m'}, \cdots, X_{m'-n+1})p_n + g_{m'}(X_{m'}, \cdots, X_1)\left[1-\sum_{n=0}^{m} p_n\right]\right)$$

注 2　如果在 N 超过 m 的条件下，模拟 N 的值有困难，则我们尝试给出 $E[g_N(X_1, X_2, \cdots, X_N)|N>m]P\{N>m\}$ 的界. 然后，选取适当大的 m 使这个界小到可以忽略. (例如，若 g_n 是取值 0 或 1 的示性函数，则 $E[g_N(X_1, X_2, \cdots, X_N)|N>m]P\{N>m\} \leqslant P\{N>m\}$，) 模拟中忽略 $E[g_N(X_1, X_2, \cdots, X_N)|N>m]P\{N>m\}$ 这项，结果仍相当精确.

8.6　重要抽样法

用 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 表示联合密度函数 (或离散时联合分布列) 是 $f(\boldsymbol{x}) = f(x_1, \cdots, x_n)$ 的随机向量，假设我们感兴趣的是估计

$$\theta = E[h(\boldsymbol{X})] = \int h(\boldsymbol{x})f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

上式是关于 $\boldsymbol{x}$ 的所有可能取值的 n 重积分（如果 X_i 是离散的，将积分看成是求 n 重和.）

由于要计算 $h(\boldsymbol{X})$ 的值，故假设直接模拟随机向量 $\boldsymbol{X}$ 是效率不高的，其可能的原因为：(a) 模拟密度函数是 $f(\boldsymbol{x}) = f(x_1, \cdots, x_n)$ 的随机向量有困难；(b) $h(\boldsymbol{X})$ 的方差太大；(c) 上述 (a) 和 (b) 的困难都存在.

下面给出通过模拟估计 θ 的另一种方法. 注意到如果 $g(\boldsymbol{x})$ 是另一个概率密度，并且当 $g(\boldsymbol{x}) = 0$ 时，有 $f(\boldsymbol{x}) = 0$，则可以将 θ 表示如下：

$$\begin{aligned}\theta &= \int \frac{h(\boldsymbol{x})f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &= E_g\left[\frac{h(\boldsymbol{X})f(\boldsymbol{X})}{g(\boldsymbol{X})}\right]\end{aligned} \tag{8.12}$$

其中 E_g 是为了强调随机向量 $\boldsymbol{X}$ 的联合密度是 $g(\boldsymbol{x})$.

从等式 (8.12) 可知，可以通过逐次的产生密度是 $g(\boldsymbol{x})$ 的随机向量 $\boldsymbol{X}$ 的值估计 θ，并且应用 $h(\boldsymbol{X})f(\boldsymbol{X})/g(\boldsymbol{X})$ 的值的平均作为 θ 的估计. 如果选取密度函数 $g(\boldsymbol{x})$ 可以使得随机变量 $h(\boldsymbol{X})f(\boldsymbol{X})/g(\boldsymbol{X})$ 的方差较小，则可以用这种方法得到 θ 的有效估计，我们称这种方法为**重要抽样法**.

下面我们体会一下重要抽样方法为什么有用. 首先，注意到当 $\boldsymbol{X}$ 是一个密度分别为 f 和 g 的随机向量时，$f(\boldsymbol{X})$ 和 $g(\boldsymbol{X})$ 分别表示给定向量 $\boldsymbol{X}$ 时各自的似然. 因此，如果 $\boldsymbol{X}$ 依照 g 分布，则通常情况下，相对于 $g(\boldsymbol{X})$ 而言 $f(\boldsymbol{X})$ 将较小，这样当依照 g 模拟出 $\boldsymbol{X}$ 时，似然比 $f(\boldsymbol{X})/g(\boldsymbol{X})$ 通常比 1 小. 但是，易知它的均值是 1： [184]

$$E_g\left[\frac{f(\boldsymbol{X})}{g(\boldsymbol{X})}\right] = \int \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \int f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$$

由此可知，尽管似然比 $f(\boldsymbol{X})/g(\boldsymbol{X})$ 通常比 1 小，但它的均值等于 1，这说明在有些时候要大于 1，因此趋向于有大的方差. 怎样才能使 $h(\boldsymbol{X})f(\boldsymbol{X})/g(\boldsymbol{X})$ 的方差较小呢？答案如下：有时我们可以适当地选取密度 g 使 $f(\boldsymbol{x})/g(\boldsymbol{x})$ 较大的那些 $\boldsymbol{x}$ 恰好对应非常小的 $h(\boldsymbol{x})$，这样比值 $h(\boldsymbol{X})f(\boldsymbol{X})/g(\boldsymbol{X})$ 通常较小. 由于这要求 $h(\boldsymbol{x})$ 有时要小，故当估计小概率时，重要抽样法看起来有最优的效果，因为在这种情况下，当 $\boldsymbol{x}$ 属于某集合时，函数 $h(\boldsymbol{x})$ 等于 1，否则，$h(\boldsymbol{x})$ 等于 0.

现在考虑怎样选取合适的密度 g. 我们发现可以使用被称作倾斜密度的函数. 令 $M(t) = E_f[\mathrm{e}^{tX}] = \int \mathrm{e}^{tx} f(x)\mathrm{d}x$ 为对应一维密度 f 的矩母函数.

定义 *密度函数*

$$f_t(x) = \frac{\mathrm{e}^{tx} f(x)}{M(t)}$$

叫做 f 的倾斜密度，$-\infty < t < \infty$.

当 $t>0$ 时，密度是 f_t 的随机变量趋向于比密度是 f 的随机变量大；当 $t<0$ 时，趋向于比密度是 f 的随机变量小.

在某些情况下，倾斜密度 f_t 与 f 有同样的参数形式.

例 8t 如果 f 是一个参数为 λ 的指数密度函数，则

$$f_t(x) = C\mathrm{e}^{tx}\lambda\mathrm{e}^{-\lambda x} = C\mathrm{e}^{-(\lambda-t)x}$$

其中 $C=1/M(t)$ 不依赖于 x. 因此，当 $t<\lambda$ 时，f_t 是一个参数为 $\lambda-t$ 的指数密度函数.

如果 f 是伯努利概率分布函数，参数是 p，则

185

$$f(x) = p^x(1-p)^{1-x}, \quad x=0,1$$

此时，$M(t)=E_f[\mathrm{e}^{tX}]=p\mathrm{e}^t+1-p$，于是有

$$\begin{aligned} f_t(x) &= \frac{1}{M(t)}(p\mathrm{e}^t)^x(1-p)^{1-x} \\ &= \left(\frac{p\mathrm{e}^t}{p\mathrm{e}^t+1-p}\right)^x\left(\frac{1-p}{p\mathrm{e}^t+1-p}\right)^{1-x} \end{aligned}$$

也就是说，f_t 是伯努利随机变量的概率分布函数，参数为 $p_t=(p\mathrm{e}^t)/(p\mathrm{e}^t+1-p)$.

如果 f 是正态密度函数，参数是 μ 和 σ^2，则 f_t 也是正态密度函数，均值是 $\mu+\sigma^2 t$，方差是 σ^2，其证明留作习题. □

在某些情况下，我们感兴趣的是某些独立随机变量 $X_1,\cdots,X_n$ 的和. 这时，联合密度 f 是一维密度的乘积，即

$$f(x_1,\cdots,x_n) = f_1(x_1)\cdots f_n(x_n)$$

其中 f_i 是 X_i 的密度函数. 在这种情况下，我们通常通过选取共同的 t，并根据它们的倾斜密度生成 X_i.

例 8u 若 $X_1,\cdots,X_n$ 是独立的随机变量，对应的概率密度 (分布) 函数分别是 f_i，$i=1,\cdots,n$. 假设感兴趣的是求它们的和至少是 a 的概率的近似值，其中 a 远大于其和的均值. 也就是说，我们感兴趣的是

$$\theta = P\{S\geqslant a\}$$

其中 $S=\sum_{i=1}^n X_i$，$a>\sum_{i=1}^n E[X_i]$. 当 $S\geqslant a$ 时，令 $I\{S\geqslant a\}$ 等于 1；否则，令 $I\{S\geqslant a\}$ 等于 0. 可得

$$\theta = E_f[I\{S\geqslant a\}]$$

其中 $\boldsymbol{f}=(f_1,\cdots,f_n)$. 假设我们根据倾斜密度函数 $f_{i,t}, i=1,\cdots,n$ 模拟 X_i ($t>0$, 其具体取值待定), 则 θ 的重要抽样估计将是:

$$\widehat{\theta}=I\{S\geqslant a\}\prod\frac{f_i(X_i)}{f_{i,t}(X_i)}$$

由于

$$\frac{f_i(X_i)}{f_{i,t}(X_i)}=M_i(t)\mathrm{e}^{-tX_i}$$

186

于是

$$\widehat{\theta}=I\{S\geqslant a\}M(t)\mathrm{e}^{-tS}$$

其中 $M(t)=\prod M_i(t)$ 是 S 的矩母函数. 由 $t>0$ 及当 $S<a$ 时, $I\{S\geqslant a\}$ 等于 0, 可得

$$I\{S\geqslant a\}\mathrm{e}^{-tS}\leqslant \mathrm{e}^{-ta}$$

于是

$$\widehat{\theta}\leqslant M(t)\mathrm{e}^{-ta}$$

为了使估计的界尽可能地小, 我们选取使 $M(t)\mathrm{e}^{-ta}$ 达到最小的 t ($t>0$). 为达到此目的, 我们给出一个估计, 在每次迭代中其值在 0 与 $\min_t M(t)\mathrm{e}^{-ta}$ 之间. 可以证明最小的 t (称之为 t^*) 满足

$$E_{t^*}[S]=E_{t^*}\left[\sum_{i=1}^{n}X_i\right]=a$$

其中的期望是在 X_i 的分布函数为 $f_{i,t^*}, i=1,\cdots,n$ 的假设下求的.

例如, 假设 $X_1,\cdots,X_n$ 是独立的伯努利随机变量, 各自的参数分别是 p_i, $i=1,\cdots,n$. 如果我们根据它们的倾斜分布函数 $p_{i,t}$, $i=1,\cdots,n$ 生成 X_i, 则 $\theta=P\{S\geqslant a\}$ 的重要抽样估计是:

$$\widehat{\theta}=I\{S\geqslant a\}\mathrm{e}^{-tS}\prod_{i=1}^{n}(p_i\mathrm{e}^t+1-p_i)$$

因为 $p_{i,t}$ 是参数为 $(p_i\mathrm{e}^t)/(p_i\mathrm{e}^t+1-p_i)$ 的伯努利随机变量的分布函数, 所以有

$$E_t\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}\frac{p_i\mathrm{e}^t}{p_i\mathrm{e}^t+1-p_i}$$

使上式等于 a 的 t 值可以通过近似计算求出并在模拟中使用.

举例说明如下：假设 $n=20, p_i=0.4, a=16$，则

187

$$E_t[S]=20\frac{0.4e^t}{0.4e^t+0.6}$$

令其等于 16，通过简单的计算可得

$$e^{t^*}=6$$

因此，如果应用参数 $(0.4e^{t^*})/(0.4e^{t^*}+0.6)=0.8$ 生成伯努利随机变量，由

$$M(t^*)=(0.4e^{t^*}+0.6)^{20} \quad 和 \quad e^{-t^*S}=(1/6)^S$$

得到重要抽样估计

$$\widehat{\theta}=I\{S\geqslant 16\}(1/6)^S 3^{20}$$

应用前面的结果得

$$\widehat{\theta}\leqslant (1/6)^{16}3^{20}=81/2^{16}=0.001\,236$$

也就是说，在每次重复中，估计值在 0 与 0.001 236 之间. 因在这种情况下，θ 是服从参数为 (20, 0.4) 的二项分布的随机变量不小于 16 的概率，故可以准确地计算出 $\theta=0.000\,317$. 因此，原来的模拟估计 I（在每次重复中，参数为 0.4 的伯努利随机变量的和小于 16 时，取值 0，否则是 1）的方差是

$$\mathrm{Var}(I)=\theta(1-\theta)=3.169\times 10^{-4}$$

另一方面，从 $0\leqslant\theta\leqslant 0.001\,236$ 可得 (见习题 29)，

$$\mathrm{Var}(\widehat{\theta})\leqslant 2.9131\times 10^{-7}$$

□

例 8v　考虑单服务员的排队系统，顾客到来的时间间隔的密度函数是 f，服务时间的密度函数是 g. 以 D_n 记第 n 个到达的顾客排队等候的时间，并假设我们感兴趣的是估计 $\alpha=P\{D_n\geqslant a\}$，其中 a 远大于 $E[D_n]$. 不是根据 f 和 g 分别产生顾客到达的时间间隔和服务时间，而应该根据 f_{-t} 和 g_t 产生这些数据，其中 t 是给定的正数. 注意到相对于用 f 和 g 产生数据，用这些分布产生的时间间隔数据较小 (因为 $-t<0$)，而服务时间数据较大. 因此，比用密度 f 和 g 模拟有更多的机会使 $D_n>a$. α 的重要抽样估计是：

188

$$\widehat{\alpha}=I\{D_n>a\}e^{t(S_n-Y_n)}[M_f(-t)M_g(t)]^n$$

其中 S_n 是前 n 个到达的时间间隔的和，Y_n 是前 n 个服务时间的和，M_f 和 M_g 分别是密度 f 和 g 的矩母函数. 所用 t 的值应根据不同的选择通过实验确定.　□

例 8w 若 $X_1, X_2, \cdots$ 是均值为 μ，方差为 1 的独立同分布的正态随机变量序列，其中 $\mu < 0$. 在质量控制中（特别是在累积和控制图分析中），一个重要的问题是确定这些值的部分和在低于 $-A$ 之前超过 B 的概率. 也就是令

$$S_n = \sum_{i=1}^{n} X_i$$

并且定义

$$N = \min\{n : S_n < -A \text{ 或者 } S_n > B\}$$

其中 A 和 B 是给定的正数. 我们感兴趣的是估计

$$\theta = P\{S_N > B\}$$

一个有效的方法是通过模拟 X_i 估计 θ. 认为它们是来自均值为 $-\mu$, 方差为 1 的正态分布，当它们的和超过 B 或低于 $-A$ 时停止，并重复该过程.（因为 $-\mu$ 是正的，所以到停止时，与用原始的负均值模拟产生的和相比，累积和更经常大于 B.）如果用 $X_1, X_2, \cdots, X_N$, 表示模拟变量（每个都是均值为 $-\mu$, 方差为 1 的正态变量），并且

$$I = \begin{cases} 1 & \text{如果} \sum_{i=1}^{N} X_i > B \\ 0 & \text{其他情况} \end{cases}$$

则从这个运行得到 θ 的模拟估计是

$$I \prod_{i=1}^{N} \left[\frac{f_\mu(X_i)}{f_{-\mu}(X_i)} \right] \tag{8.13}$$

其中 f_c 是均值为 c 方差为 1 的正态密度函数. 因为

$$\frac{f_\mu(x)}{f_{-\mu}(x)} = \frac{\exp\{-\frac{(x-\mu)^2}{2}\}}{\exp\{-\frac{(x+\mu)^2}{2}\}} = \mathrm{e}^{2\mu x}$$

189

则由 (8.13) 得到基于这个运行的 θ 的估计是

$$I\exp\left\{2\mu \sum_{i=1}^{N} X_i\right\} = I\exp\{2\mu S_N\}$$

当 I 等于 1 时，S_N 超过 B，并且由此时 $\mu < 0$ 得估计小于 $\mathrm{e}^{2\mu B}$. 也就是说，不用通过模拟运行得到值 0 或者 1（如果我们直接模拟时，将会得到这样的结果），在这种情况下，得到值 0 或者比 $\mathrm{e}^{2\mu B}$ 小的正值，有力地表明为什么重要抽样方法可以

减小方差. 例如, 如果 $\mu=-0.1$, $B=5$, 则每次运行的估计在 0 与 $e^{-1}=0.3679$ 之间. 并且这一点在理论也同样重要, 因为它表明

$$P\{\text{到达} -A \text{ 之前超过} B\} \leqslant e^{2\mu B}$$

由于上式对任何正的 A 而言都对, 则我们得到一个有意思的结论:

$$P\{\text{曾经超过} B\} \leqslant e^{2\mu B}$$

□

例 8x　若 $\boldsymbol{X}=(X_1,\cdots,X_{100})$ 是 $(1,2,\cdots,100)$ 的一个随机排列. 也就是说, $\boldsymbol{X}$ 等可能地等于 100! 个排列中的任意一个. 假设我们感兴趣的是用模拟估计

$$\theta=P\left\{\sum_{j=1}^{100} jX_j > 290\,000\right\}$$

为了对 θ 的大小有一个大致了解, 我们可以从计算 $\sum_{j=1}^{100} jX_j$ 的均值和标准差开始. 实际上, 不难证明

$$E\left[\sum_{j=1}^{100} jX_j\right]=100(101)^2/4=255\,025$$

$$SD\left(\sum_{j=1}^{100} jX_j\right)=\sqrt{(99)(100)^2(101)^2/144}=8374.478$$

因此, 如果假设 $\sum_{j=1}^{100} jX_j$ 的分布大致是正态的, 用 Z 表示服从标准正态分布的随机变量, 则有

$$\begin{aligned}\theta &\approx P\left\{Z>\frac{290\,000-255\,025}{8374.478}\right\}\\ &=P\{Z>4.1764\}\\ &=0.000\,014\,81\end{aligned}$$

190

由于 θ 显然是一个小概率, 所以考虑使用重要抽样法估计是有意义的.

为了使用重要抽样法, 我们想要产生能使 $\sum_{j=1}^{100} jX_j > 290\,000$ 的概率较大的 $\boldsymbol{X}$ 的排列. 实际上, 我们试着得到概率大约为 0.5. 当 $X_j=j, j=1,\cdots,100$ 时, $\sum_{j=1}^{100} jX_j$ 将达到最大值, 并且实际上, 在 j 较大时, 若 X_j 趋于较大, 则它的值也将趋于较大, 而在 j 较小时, 它的值也将趋于较小. 生成一个趋于这种类型的 $\boldsymbol{X}$ 的排列的方法如下: 生成参数为 λ_j 的独立的指数随机变量 $Y_j, j=1,\cdots,100$, 其

中 $\lambda_j, j=1,\cdots,100$ 是单调增的序列，具体值将在下面给出. 对于 $j=1,\cdots,100$，用 X_j 表示生成的这些值中第 j 大的数. 也就是说，

$$Y_{X_1} > Y_{X_2} > \cdots > Y_{X_{100}}$$

因此，对值大的 j，Y_j 将趋于是 Y 中值较小的，因此，当 j 较大时，X_j 的值也将趋于较大，并且因此 $\sum_{j=1}^{100} jX_j$ 的值将趋于较大，相对于 X 是均匀分布时的排列的值.

现在计算 $E[\sum_{j=1}^{100} jX_j]$. 要求这个值，用 $R(j)$ 表示 $Y_j, j=1,\cdots,100$ 的秩，这里最大数的秩是 1，第 2 大的数的秩是 2，依从此类推直到最小的数的秩是 100. 注意到因为 X_j 是 Y 中第 j 大的数的下标，于是 $R(X_j)=j$. 因此，

$$\sum_{j=1}^{100} jX_j = \sum_{j=1}^{100} R(X_j)X_j = \sum_{j=1}^{100} jR(j)$$

其中最后一个等式成立的原因是 $X_1,\cdots,X_{100}$ 是 $1,\cdots,100$ 的一个排列. 所以有

$$E\left[\sum_{j=1}^{100} jX_j\right] = \sum_{j=1}^{100} jE[R(j)]$$

为计算 $E[R_j]$，当 $Y_j < Y_i$ 时，令 $I(i,j)=1$；否则，令它为 0. 并且注意到：

$$R_j = 1 + \sum_{i:i\neq j} I(i,j)$$

总而言之，上式表明 Y_j 的秩是 1 加上比 Y_j 大的 Y_i 的个数. 利用

$$P\{Y_j < Y_i\} = \frac{\lambda_j}{\lambda_i + \lambda_j}$$

191

求期望得到

$$E[R_j] = 1 + \sum_{i:i\neq j} \frac{\lambda_j}{\lambda_i + \lambda_j}$$

如此可得

$$E\left[\sum_{j=1}^{100} jX_j\right] = \sum_{j=1}^{100} j\left(1 + \sum_{i:i\neq j} \frac{\lambda_j}{\lambda_i + \lambda_j}\right)$$

如果我们令 $\lambda_j = j^{0.7}, j=1,\cdots,100$，通过计算可得 $E[\sum_{j=1}^{100} jX_j] = 290\ 293.6$，因此，当以这些参数生成 $\boldsymbol{X}$ 时，将会有

$$P\left\{\sum_{j=1}^{100} jX_j > 290\ 000\right\} \approx 0.5$$

因此，我们认为模拟估计应通过以下步骤得到：先以各自的速率 $j^{0.7}$ 产生独立指数变量 Y_j，再用 X_j 作为第 j 个大的值的下标，$j=1,\cdots,100$. 当 $\sum_{j=1}^{100} jX_j > 290\,000$ 时，令 $I=1$；否则，令它为 0. 这样，当 $Y_{X_{100}}$ 是最小的 Y，$Y_{X_{99}}$ 是第二小的 Y……时，将得到输出的结果 $\boldsymbol{X}$. 当 $\boldsymbol{X}$ 等可能地等于排列中的任何一个时，这个结果的概率将是 $1/(100!)$，尽管按上述方式进行模拟时，概率为

$$\frac{(X_{100})^{0.7}}{\sum_{j=1}^{100}(X_j)^{0.7}}\frac{(X_{99})^{0.7}}{\sum_{j=1}^{99}(X_j)^{0.7}}\cdots\frac{(X_2)^{0.7}}{\sum_{j=1}^{2}(X_j)^{0.7}}\frac{(X_1)^{0.7}}{(X_1)^{0.7}}$$

因此，从单个运行得到的重要抽样估计为：

$$\widehat{\theta}=\frac{I}{(100)!}\frac{\prod_{j=1}^{100}\left(\sum_{n=1}^{n}(X_j)^{0.7}\right)}{\left(\prod_{n=1}^{100}n\right)^{0.7}}=\frac{I\prod_{n=1}^{100}\left(\sum_{j=1}^{n}(X_j)^{0.7}\right)}{\left(\prod_{n=1}^{100}n\right)^{1.7}}$$

在开始模拟之前，应先计算出 $C=1.7\sum_{n=1}^{100}\ln(n)$ 和 $a(j)=-j^{-0.7}, j=1,\cdots,100$ 的值. 按以下的步骤进行模拟：

对 $j=1$ 到 100,
分别产生随机数 U,
令 $Y_j=a(j)\ln U$,
接着,
对 $X_j, j=1,\cdots,100$, 令 Y_{X_j} 是第 j 大的 Y,
如果 $\sum_{j=1}^{n} jX_j \leqslant 290\,000$, 令 $\widehat{\theta}=0$ 并停止,
192
令 $S=0, P=0$,
对 $n=1$ 到 100,
$S=S+(X_n)^{0.7}$,
$P=P+\ln S$,
接着,
令 $\widehat{\theta}=\mathrm{e}^{P-C}$.

用 50 000 次模拟运行得到估计 $\widehat{\theta}=3.77\times10^{-6}$，其样本方差等于 1.89×10^{-8}. 由于原始的模拟估计（当 $\sum_{j=1}^{100} jX_j > 290\,000$ 时，$I=1$；否则，它为 0）的方差 $\mathrm{Var}(I)=\theta(1-\theta)\approx3.77\times10^{-6}$，于是有：

$$\frac{\mathrm{Var}(I)}{\mathrm{Var}(\widehat{\theta})}\approx199.47$$

□

我们得到，当条件是小概率事件时，重要抽样方法在估计条件期望时也是十分有用的. 也就是说，假设 $\boldsymbol{X}$ 是一个密度为 f 的随机向量，我们感兴趣的是估计

$$\theta = E[h(\boldsymbol{X})|\boldsymbol{X} \in \mathcal{A}]$$

其中 $h(\boldsymbol{x})$ 是任意实值函数，$P\{\boldsymbol{X} \in \mathcal{A}\}$ 是一个未知的小概率. 由于在给定 $\boldsymbol{X}$ 属于 $\mathcal{A}$ 时，$\boldsymbol{X}$ 的条件密度是

$$f(\boldsymbol{x}|\boldsymbol{X} \in \mathcal{A}) = \frac{f(\boldsymbol{x})}{P\{\boldsymbol{X} \in \mathcal{A}\}}, \quad \boldsymbol{x} \in \mathcal{A}$$

故

$$\begin{aligned}\theta &= \frac{\int_{\boldsymbol{x}\in\mathcal{A}} h(\boldsymbol{x})f(\boldsymbol{x})\mathrm{d}(\boldsymbol{x})}{P\{\boldsymbol{X} \in \mathcal{A}\}} \\ &= \frac{E[h(\boldsymbol{X})I(\boldsymbol{X} \in \mathcal{A})]}{E[I(\boldsymbol{X} \in \mathcal{A})]} \\ &= \frac{E[N]}{E[D]}\end{aligned}$$

其中 $E[N]$ 和 $E[D]$ 分别等于上式的分子和分母，$I(\boldsymbol{X} \in \mathcal{A})$ 定义为：当 $\boldsymbol{X} \in \mathcal{A}$ 时，等于 1；否则，等于 0. 因此，我们不是根据密度 f 模拟 $\boldsymbol{X}$，因为这样使模拟结果属于 $\mathcal{A}$ 不太可能；而是根据另一个密度 g 进行模拟，因为这样模拟结果属于 $\mathcal{A}$ 的可能性更大. 如果根据 g 模拟出 k 个随机向量 $\boldsymbol{X}^1, \cdots, \boldsymbol{X}^k$，则我们可通过 $\frac{1}{k}\sum_{i=1}^k N_i$ 估计 $E[N]$，通过 $\frac{1}{k}\sum_{i=1}^k D_i$ 估计 $E[D]$，其中

$$N_i = \frac{h(\boldsymbol{X}^i)I(\boldsymbol{X}^i \in \mathcal{A})f(\boldsymbol{X}^i)}{g(\boldsymbol{X}^i)}$$

193

和

$$D_i = \frac{I(\boldsymbol{X}^i \in \mathcal{A})f(\boldsymbol{X}^i)}{g(\boldsymbol{X}^i)}$$

于是，我们得到 θ 的如下形式估计：

$$\hat{\theta} = \frac{\sum_{i=1}^{k} h(\boldsymbol{X}^i)I(\boldsymbol{X}^i \in \mathcal{A})f(\boldsymbol{X}^i)/g(\boldsymbol{X}^i)}{\sum_{i=1}^{k} I(\boldsymbol{X}^i \in \mathcal{A})f(\boldsymbol{X}^i)/g(\boldsymbol{X}^i)} \tag{8.14}$$

这个估计的均方误差可以用自助法进行估计（见例 7e）.

例 8y 若 X_i 是独立的服从指数分布的随机变量，参数分别为 $1/(i+2), i = 1, 2, 3, 4$. 令 $S = \sum_{i=1}^4 X_i$，并假设我们要估计 $\theta = E[S|S > 62]$. 也就是说，我们可

以先选取一个 t 的值，然后模拟参数等于 $1/(i+2)-t$ 的 X_i. 如果选取 $t=0.14$，则 $E_t[S]=68.43$. 因此，我们生成 k 个参数等于 $1/(i+2)-0.14$ 的指数随机变量 X_i 的集合，$i=1,2,3,4$，并令 S_j 是第 j 个集合的和，$j=1,\cdots,k$. 于是可通过

$$\frac{C}{k}\sum_{j=1}^{k} S_j I(S_j>62)\mathrm{e}^{-0.14S_j} \quad 估计 \quad E[SI(S>62)]$$

$$\frac{C}{k}\sum_{j=1}^{k} I(S_j>62)\mathrm{e}^{-0.14S_j} \quad 估计 \quad E[I(S>62)]$$

其中 $C=\prod_{i=1}^{4}\frac{1}{1-0.14(i+2)}=81.635$. θ 的估计是

$$\hat{\theta}=\frac{\sum\limits_{j=1}^{k} S_j I(S_j>62)\mathrm{e}^{-0.14S_j}}{\sum\limits_{j=1}^{k} I(S_j>62)\mathrm{e}^{-0.14S_j}}$$

□

我们也可以用重要抽样方法只通过一个模拟而估计两个（或多个）不同量. 例如，假设

$$\theta_1=E[h(\boldsymbol{Y})] \quad \theta_2=E[h(\boldsymbol{W})]$$

其中 $\boldsymbol{Y}$ 和 $\boldsymbol{W}$ 是随机向量，各自的联合密度函数分别是 f 和 g. 如果模拟出 $\boldsymbol{W}$，[194] 则我们可以在用 $h(\boldsymbol{W})$ 估计 θ_2 的同时用 $h(\boldsymbol{W})f(\boldsymbol{W})/g(\boldsymbol{W})$ 估计 θ_1. 例如，假设模拟 T，即在一个排队系统中前 r 个到达的顾客花费在系统中时间的总和，假设顾客接受服务时间服从均值是 2 的指数分布. 如果我们发现在这个系统中顾客接受服务时间应该是参数为 $(2,1)$ 的伽玛分布，则并不需要进行重新模拟，我们可以应用估计

$$T\frac{\prod\limits_{i=1}^{r} S_i\exp\{-S_i\}}{\prod\limits_{i=1}^{r}\left(\frac{1}{2}\exp\{-S_i/2\}\right)}=2^r T\exp\left\{-\sum_{i=1}^{r}\frac{S_i}{2}\right\}\prod_{i=1}^{r}S_i$$

其中 S_i 是产生（指数型）的第 i 个顾客接受服务的时间.[可以使用上述方法是因为指数型的顾客接受服务的时间的密度是 $g(s)=\frac{1}{2}\mathrm{e}^{-s/2}$，而参数为 $(2,1)$ 的伽玛分布的密度是 $f(s)=s\mathrm{e}^{-s}$.]

如随机变量 X 的密度 f 虽然已知，但其分布函数的值难以估计，则我们可以利用重要抽样方法来估计此随机变量 X 的尾部概率. 假设我们要估计 $P_f\{X>a\}$，其中下标 f 用来表明 X 的密度函数是 f，a 的值是给定的. 令

$$I(X>a)=\begin{cases}1 & \text{如果}X>a\\ 0 & \text{如果}X\leqslant a\end{cases}$$

则有

$$\begin{aligned}P_f\{X>a\}&=E_f[I(X>a)]\\&=E_g\left[I(X>a)\frac{f(X)}{g(X)}\right] \quad \text{重要抽样等式}\\&=E_g\left[I(X>a)\frac{f(X)}{g(X)}\bigg|X>a\right]P_g\{X>a\}\\&\quad+E_g\left[I(X>a)\frac{f(X)}{g(X)}\bigg|X\leqslant a\right]P_g\{X\leqslant a\}\\&=E_g\left[\frac{f(X)}{g(X)}\bigg|X>a\right]P_g\{X>a\}\end{aligned}$$

如果 g 是指数密度

$$g(x)=\lambda\mathrm{e}^{-\lambda x},\ x>0$$

则上述表明，对于 $a>0$,

$$P_f\{X>a\}=\frac{\mathrm{e}^{-\lambda a}}{\lambda}E_g[\mathrm{e}^{\lambda X}f(X)|X>a]$$

195

因为指数随机变量在大于 a 时的条件分布与指数随机变量本身加上 a 的分布相同，所以由上可得

$$\begin{aligned}P_f\{X>a\}&=\frac{\mathrm{e}^{-\lambda a}}{\lambda}E_g[\mathrm{e}^{\lambda(X+a)}f(X+a)]\\&=\frac{1}{\lambda}E_g[\mathrm{e}^{\lambda X}f(X+a)]\end{aligned}$$

因此，可以通过产生参数为 λ 的独立的指数随机变量 $X_1,\cdots,X_k$, 估计尾部概率 $P_f\{X>a\}$, 并用

$$\frac{1}{\lambda}\frac{1}{k}\sum_{i=1}^{k}\mathrm{e}^{\lambda X_i}f(X_i+a)$$

作为估计.

作为上面方法的应用，假设 f 是标准正态分布随机变量 Z 的密度函数，并且 $a>0$. 若 X 是参数为 λ 的指数随机变量，则由上可得

$$\begin{aligned}P\{Z>a\}&=\frac{1}{a\sqrt{2\pi}}E[\mathrm{e}^{aX-(X+a)^2/2}]\\&=\frac{\mathrm{e}^{-a^2/2}}{a\sqrt{2\pi}}E[\mathrm{e}^{-X^2/2}]\end{aligned}$$

因此，我们可以通过产生参数为 a 的指数随机变量 X 估计 $P\{Z>a\}$，并用

$$EST=\frac{\mathrm{e}^{-a^2/2}}{a\sqrt{2\pi}}\mathrm{e}^{-X^2/2}$$

作为估计. 为了计算这个估计的方差，注意到

$$\begin{aligned}E[\mathrm{e}^{-X^2/2}]&=\int_0^\infty \mathrm{e}^{-x^2/2}a\mathrm{e}^{-ax}\mathrm{d}x\\&=a\int_0^\infty \exp\{-(x^2+2ax)/2\}\mathrm{d}x\\&=a\mathrm{e}^{a^2/2}\int_0^\infty \exp\{-(x+a)^2/2\}\mathrm{d}x\\&=a\mathrm{e}^{a^2/2}\int_a^\infty \exp\{-y^2/2\}\mathrm{d}y\\&=a\mathrm{e}^{a^2/2}\sqrt{2\pi}\ \bar{\Phi}\ (a)\end{aligned}$$

196

通过类似讨论，可得

$$E[\mathrm{e}^{-X^2}]=a\mathrm{e}^{a^2/4}\sqrt{\pi}\ \bar{\Phi}\ (a/\sqrt{2})$$

结合上面的论述，可得到 $\mathrm{Var}(EST)$. 例如，当 $a=3$ 时

$$E[\mathrm{e}^{-X^2/2}]=3\mathrm{e}^{4.5}\sqrt{2\pi}\ \bar{\Phi}\ (3)\approx 0.9138$$

和

$$E[\mathrm{e}^{-X^2}]=3\mathrm{e}^{2.25}\sqrt{\pi}\ \bar{\Phi}\ (2.1213)\approx 0.8551$$

于是

$$\mathrm{Var}(\mathrm{e}^{-X^2/2})\approx 0.8551-(0.9138)^2=0.0201$$

因为 $\frac{\mathrm{e}^{-4.5}}{3\sqrt{2\pi}}\approx 0.001\ 477$，故当 $a=3$ 时，我们给出

$$\mathrm{Var}(EST)=(0.001\ 477)^2\mathrm{Var}(\mathrm{e}^{-X^2/2})\approx 4.38\times 10^{-8}$$

作为比较，原始的模拟估计（当产生的标准正态随机变量值超过 3 时值为 1，否则值为 0）的方差是 $P\{Z>3\}(1-P\{Z>3\})\approx 0.001\ 34$. 实际上，ETS 的方差是相当地小，在置信度为 95% 时，从单个指数模拟得到的估计与准确值的偏差在 ±0.0004 之内.

8.7　公共随机数的应用

假设有 n 件工作可由一对相同机器中的任一个进行处理. 用 T_j 表示处理第 j 个工作所用的时间，$j=1,\cdots,n$. 我们感兴趣的是在两种不同工作安排顺序的策略

下，对处理完所有工作所需的时间进行比较，第一种叫长工期工作优先策略，即当一个机器有空闲时，选取剩余的工作中需要处理时间最长的先来处理，第二种叫短工期工作优先策略，即选取剩余的工作中需要处理时间最短的先来处理. 例如，如果 $n=3, T_1=2, T_2=5, T_3=3$，则采用长工期工作优先策略完成任务需时间数是 5，而采用短工期工作优先策略需时间数是 7. 当处理工作的时间 $T_1,\cdots,T_n$ 是分布为 F 的随机变量时，我们将通过模拟比较在这两种不同策略下，完工时间期望的差异.

197

换句话说，如果 $g(t_1,\cdots,t_n)$ 是采用长工期工作优先策略下，处理需要时间分别为 $t_1,\cdots,t_n$ 的 n 个工作所用的时间，而 $h(t_1,\cdots,t_n)$ 是采用短工期工作优先策略下，处理这 n 个工作所用的时间，则我们感兴趣的是用模拟估计

$$\theta=\theta_1-\theta_2$$

其中

$$\theta_1=E[g(\boldsymbol{T})],\quad \theta_2=E[h(\boldsymbol{T})],\quad \boldsymbol{T}=(T_1,\cdots,T_n)$$

如果现在我们通过产生向量 $\boldsymbol{T}$ 来计算 $g(\boldsymbol{T})$，则接下来的问题是：应该用这些同样的值计算 $h(\boldsymbol{T})$，还是重新产生一组独立的值估计 θ_2 更有效. 为了回答这个问题，假设我们用与 $\boldsymbol{T}$ 的分布相同的 $\boldsymbol{T}^*=(T_1^*,\cdots,T_n^*)$ 估计 θ_2. 则 θ 的估计 $g(\boldsymbol{T})-h(\boldsymbol{T}^*)$ 的方差是

$$\begin{aligned}\operatorname{Var}(g(\boldsymbol{T})-h(\boldsymbol{T}^*))&=\operatorname{Var}(g(\boldsymbol{T}))+\operatorname{Var}(h(\boldsymbol{T}^*))-2\operatorname{Cov}(g(\boldsymbol{T}),h(\boldsymbol{T}^*))\\&=\operatorname{Var}(g(\boldsymbol{T}))+\operatorname{Var}(h(\boldsymbol{T}))-2\operatorname{Cov}(g(\boldsymbol{T}),h(\boldsymbol{T}^*))\end{aligned}\tag{8.15}$$

因此，如果 $g(\boldsymbol{T})$ 与 $h(\boldsymbol{T})$ 正相关，即如果它们的协方差是正的，则用 $g(\boldsymbol{T})$ 与 $h(\boldsymbol{T})$ 得到的 θ 的估计的方差，比用与 $\boldsymbol{T}$ 独立的 $\boldsymbol{T}^*$ 计算 $h(\boldsymbol{T}^*)$ 得到的 θ 的估计的方差小. [在后一种情况下，(8.15) 中的协方差是 0.]

由于 g 和 f 都是它们的自变量的增函数，而独立随机变量的增函数是正相关的（证明见本章附录），于是就上述情况而言，在这两种策略下，通常使用同一次产生的工作时间进行计算效果较好.

一般性标准为：当比较随机确定的环境下不同的操作规则时，在模拟出环境状态以后，应该在这个环境状态下评估所有的策略. 也就是，如果环境由向量 $\boldsymbol{T}$ 确定，并且 $g_i(\boldsymbol{T})$ 是在环境状态 $\boldsymbol{T}$ 下按策略 i 返回的值，则在模拟出随机向量 $\boldsymbol{T}$ 之后，对应 $\boldsymbol{T}$ 的值，应该求出所有的返回值 $g_i(\boldsymbol{T})$.

8.8 对奇异期权的评估

以时刻 0 表示当前的时间，以 $P(y)$ 表示在时刻 y 股票的价格. 一般假设股票

[198] 价格按几何布朗运动过程随时间变化. 这意思是说, 对任何过去时刻 y 的股票价格, 时刻 $t+y$ 的价格与时刻 y 的价格之比服从均值为 μt, 方差为 $t\sigma^2$ 的对数正态分布. 即随机变量

$$\ln\left(\frac{P(t+y)}{P(y)}\right)$$

与时刻 y 的价格独立, 且服从均值为 μt, 方差为 $t\sigma^2$ 的正态分布. 参数 μ 和 σ 分别叫做几何布朗运动的漂移和波动率.

对于欧式买入期权, 若其到期日是 t, 敲定价是 K, 则此期权的所有者拥有在时刻 t 以固定价格 K 购买股票的权力 (并不是义务). 若在时刻 t 股价 $P(t)>K$, 则将执行期权, 即以价格 K 买进市场价是 $P(t)$ 的股票, 且在这个交易中的收益是 $P(t)-K$. 因此, 一般来说, 该期权在时刻 t 的收益是

$$(P(t)-K)^+$$

其中

$$x^+=\begin{cases} x & \text{如果 } x>0 \\ 0 & \text{如果 } x\leqslant 0 \end{cases}$$

对给定的初始价格 $P(0)=v$, 以 $C(K,t,v)$ 记 K,t 欧式买入期权的期望收益. 由

$$W\equiv\ln(P(t)/v)$$

是一个均值为 $t\mu$, 方差为 $t\sigma^2$ 的正态随机变量, 可得

$$C(K,t,v)=E[(P(t)-K)^+]=E[(v\mathrm{e}^W-K)^+]$$

通过对上式的直接计算不难得到 $C(K,t,v)$.

上面的期权被称作常规 (或普通) 的买入期权. 近年来, 人们的目光主要集中在非常规 (或奇异) 期权方面. 障碍期权 (barrier option) 就是一种非常规期权; 当资产价格达到某个障碍水平时, 该期权得以存在或作废. 我们将考虑一种叫上升敲入 (up-and-in) 的障碍期权, 它不但受控于价格 K 和时刻 t, 而且受控于附加价格 b 和附加时刻 $s, s<t$. 这种期权的交易是这样的: 只有在时刻 s, 股价超过 b 时, 它 [199] 的持有者才有在时刻 t 以价格 K 购买的权力. 换句话说, 如果 $P(s)>b$, 则在时刻 s 该 K,t 期权得以存在; 如果 $P(s)\leqslant b$, 则该 K,t 期权作废. 我们来看怎样有效地运用模拟给出这样一个期权的期望收益.

假设 $P(0)=v$, 定义 X 和 Y 如下:

$$X=\ln\left(\frac{P(s)}{v}\right),\quad Y=\ln\left(\frac{P(t)}{P(s)}\right)$$

由几何布朗运动的性质知，X 和 Y 是相互独立的正态随机变量，X 的均值是 $s\mu$，方差是 $s\sigma^2$，Y 的均值是 $(t-s)\mu$，方差是 $(t-s)\sigma^2$. 因为

$$P(s) = v\mathrm{e}^X$$

$$P(t) = v\mathrm{e}^{X+Y}$$

则可以将期权的收益写为

$$\text{收益} = I(v\mathrm{e}^X > b)(v\mathrm{e}^{X+Y} - K)^+$$

其中

$$I(v\mathrm{e}^X > b) = \begin{cases} 1 & \text{如果 } v\mathrm{e}^X > b \\ 0 & \text{如果 } v\mathrm{e}^X \leqslant b \end{cases}$$

因此，可以通过产生一对正态随机变量对收益进行模拟.

原始的模拟估计是先产生 X. 如果 X 小于 $\ln(b/v)$，则以收益为 0 运行结束；如果 X 大于 $\ln(b/v)$，则产生 Y，并且该运行收益为 $(v\mathrm{e}^{X+Y} - K)^+$.

然而，通过联合使用分层抽样和条件期望这两种缩减方差的方法可以显著地改进模拟的效果. 为达到改进的目的，以 R 记期权的收益，并有

$$\begin{aligned} E[R] &= E[R|v\mathrm{e}^X > b]P\{v\mathrm{e}^X > b\} + E[R|v\mathrm{e}^X \leqslant b]P\{v\mathrm{e}^X \leqslant b\} \\ &= E[R|X > \ln(b/v)]P\{X > \ln(b/v)\} \\ &= E[R|X > \ln(b/v)]\bar{\Phi}\left(\frac{\ln(b/v) - s\mu}{\sigma\sqrt{s}}\right) \end{aligned}$$

其中 $\bar{\Phi} = 1 - \Phi$ 是标准正态分布的尾部概率分布函数. 因此，为得到 $E[R]$，我们仅需确定在给定 $X > \ln(b/v)$ 下它的条件期望即可，而此条件期望的求得可以通过在其大于 $\ln(b/v)$ 的条件下产生 X 的随机数来完成. 假设产生的值是 x（随后我们将给出怎样产生以超过某一值为条件的正态数据）. 现在，我们不是使用产生的 Y 值确定模拟的收益，而是使用在给定 X 的值的条件下收益的条件期望作为估计. 这个条件期望可以计算出的原因是：当 $X > \ln(b/v)$ 时，在时刻 s 该期权得以存在，因此与常规期权有相同的期望收益，此时该证券的初始价格是 $v\mathrm{e}^X$，$t-s$ 之后到期. 也就是，在以大于 $\ln(b/v)$ 为条件模拟出 X 之后，我们将用如下的估计作为障碍期权的期望收益： 200

$$\text{估计} = C(K, t-s, v\mathrm{e}^X)\bar{\Phi}\left(\frac{\ln(b/v) - s\mu}{\sigma\sqrt{s}}\right) \tag{8.16}$$

在模拟 k 次之后，以 X_i 记第 i 次模拟产生的条件正态分布的值，则所得估计是

$$\bar{\Phi}\left(\frac{\ln(b/v)-s\mu}{\sigma\sqrt{s}}\right)\frac{1}{k}\sum_{i=1}^{k}C(K,t-s,v\mathrm{e}^{X_i})$$

下面给出在 X 大于 $\ln(b/v)$ 条件下如何产生 X 的方法. 因为 X 可表示为

$$X=s\mu+\sigma\sqrt{s}Z \tag{8.17}$$

其中 Z 是服从标准正态分布的随机变量，这等价于以事件

$$Z>c\equiv\frac{\ln(b/v)-s\mu}{\sigma\sqrt{s}} \tag{8.18}$$

为条件产生 Z. 因此，我们需要产生以大于 c 为条件的服从标准正态分布的随机变量. 当 $c\leqslant 0$ 时，我们可以直接产生服从标准正态分布的随机变量，直到有一个大于 c. 更有趣的情形是当 $c>0$ 时，在这种情况下，一个更有效的程序是对 $c+Y$ 的密度函数 g 应用筛选法，其中 Y 是指数随机变量，参数 λ 在后面给出. $c+Y$ 的密度函数是

$$g(x)=\lambda\mathrm{e}^{-\lambda x}\mathrm{e}^{\lambda c}=\lambda\mathrm{e}^{-\lambda(x-c)},\quad x>c$$

考虑到大于 c 的标准正态分布的条件密度是

201

$$f(x)=\frac{1}{\sqrt{2\pi}\bar{\Phi}(c)}\mathrm{e}^{-x^2/2},\quad x>c$$

从而

$$\frac{f(x)}{g(x)}=\frac{\mathrm{e}^{-\lambda c}\mathrm{e}^{\lambda x-x^2/2}}{\lambda\bar{\Phi}(c)\sqrt{2\pi}}$$

因为当 $x=\lambda$ 时, $\mathrm{e}^{\lambda x-x^2/2}$ 取最大值, 可得

$$\max_x\frac{f(x)}{g(x)}\leqslant C(\lambda)\equiv\frac{\mathrm{e}^{\lambda^2/2-\lambda c}}{\lambda\bar{\Phi}(c)\sqrt{2\pi}}$$

计算表明当

$$\lambda=\frac{c+\sqrt{c^2+4}}{2}$$

时, $C(\lambda)$ 取最小值. 由上式可得 λ 的值. 因为

$$\frac{f(x)}{C(\lambda)g(x)}=\mathrm{e}^{\lambda x-x^2/2-\lambda^2/2}=\mathrm{e}^{-(x-\lambda)^2/2}$$

则我们可用如下的算法来产生在取值大于正值 c 的条件下的标准正态随机变量.

1. 令 $\lambda = \frac{c+\sqrt{c^2+4}}{2}$.
2. 产生 U_1, 并令 $Y = -\frac{1}{\lambda}\ln(U_1)$ 和 $V = c + Y$.
3. 产生 U_2.
4. 如果 $U_2 \leqslant \mathrm{e}^{-(V-\lambda)^2/2}$ 则停止; 否则回到 2.

所得 V 值的分布是在取值大于 $c(>0)$ 的条件下的标准正态分布.

注 1 上面使用的产生在取值大于 c 的条件下的标准正态随机变量的运算方法十分有效, 特别是当 c 较大时. 例如, 如果 $c = 3$, 则有 $\lambda \approx 3.3$ 和 $C(\lambda) \approx 1.04$.

注 2 第 4 步的不等式可改写为

$$-\ln(U_2) \geqslant (V-\lambda)^2/2$$

注意 $-\ln(U_2)$ 是参数为 1 的指数随机变量, 而取值超过某一给定值的指数随机变量仍是同参数的指数随机变量. 于是, 上述算法不但产生了取值大于 c 的条件下的标准正态随机变量, 而且也得到了一个独立的参数为 1 的指数随机变量, 它可以用来产生下一个条件标准正态随机变量.

202

注 3 由于常规期权的期望收益 $C(K, t, v)$ 是股票初始价格 v 的增函数, 故由 (8.16) 式给出的估计在 X 上也是递增的. 根据等式 (8.17), 等价于 (8.16) 的估计在 Z 上是递增的. 这建议我们可以将 Z 作为控制变量, 因为 Z 是以不等式 (8.18) 为条件产生的, 它的均值是:

$$\begin{aligned} E[Z|Z>c] &= \frac{1}{\sqrt{2\pi}\bar{\Phi}(c)}\int_c^{\infty} x\mathrm{e}^{-x^2/2}\mathrm{d}x \\ &= \frac{\mathrm{e}^{-c^2/2}}{\sqrt{2\pi}\bar{\Phi}(c)} \end{aligned}$$

注 4 障碍期权的期望回报可以表示成正态密度函数乘积的二重积分. 可根据二元正态随机变量的联合概率分布求这个二重积分的值. 然而, 对于比 $(P(t)-K)^+$ 更一般的收益函数, 例如 $[(P(t)-K)^+]^\alpha$ 形式的幂收益函数, 这种方法就不可用, 而上述的模拟程序则可能是估计期望收益最有效的方法. □

8.9 随机排列和随机子集的函数的估计

在这一节, 我们介绍新的方差缩减技术, 并将该技术应用于模拟估计随机排列的函数和随机子集的函数的期望值.

随机排列

令 $I_1, \cdots, I_n$ 等可能地等于 $1, \cdots, n$ 的 $n!$ 个排列中的任意一个, 假设我们感兴趣的是用模拟估计 $\theta = E[f(v_{I_1}, \cdots, v_{I_n})]$, 其中 $v_1 < v_2 < \cdots < v_n$ 且函数 f 都

是给定的. 在产生一个随机排列 $\boldsymbol{V} = (v_{I_1}, \cdots, v_{I_n})$ 之后,"逆排列" 即该排列的逆顺序排列立即可知, 即

$$\boldsymbol{V_1} = (v_{I_n}, \cdots, v_{I_1})$$

我们将证明, 如果 f 是一个可交换单调函数 (定义在后面), 则用 $\boldsymbol{V}_1$ 与 $\boldsymbol{V}$ 联合的模拟估计比用另一个与 $\boldsymbol{V}$ 独立的随机排列与 $\boldsymbol{V}$ 联合的模拟估计要好.

203

令 $\boldsymbol{V}^1 = (v_{i_1}, \cdots, v_i, \cdots, v_j, \cdots, v_{i_n})$ 与 $\boldsymbol{V}^2 = (v_{i_1}, \cdots, v_j, \cdots, v_i, \cdots, v_{i_n})$ 是除 v_i 与 v_j 交换之外的 $v_1, \cdots, v_n$ 的同一排列. 称定义在排列 $v_1, \cdots, v_n$ 的函数 f 是可交换的增函数, 如果

$$f(\boldsymbol{V^1}) \geqslant f(\boldsymbol{V^2}), \text{ 当 } v_i > v_j \text{ 时}$$

称函数 f 是可交换的减函数, 如果

$$f(\boldsymbol{V^1}) \leqslant f(\boldsymbol{V^2}), \text{ 当 } v_i > v_j \text{ 时}$$

也就是说, 当排列中较大的元素交换到前面的时候, 排列的可交换增 (减) 函数的值增加 (减少).

例 8z 假设 n 项工作必须依次在一台机器进行处理, 假设处理时间分别是 $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$, 且在时间 t 完成一项工作的报酬是 $R(t)$. 因此, 按顺序 $i_1, \cdots, i_n$ 完成工作的总报酬是

$$f(i_1, \cdots, i_n) = \sum_{j=1}^{n} R(t_{i_1} + \cdots + t_{i_j})$$

易证当 $R(t)$ 是 t 的减函数时, f 是可交换的减函数; 当 $R(t)$ 是 t 的增函数时, f 是可交换的增函数. □

在下面, 假设 $\boldsymbol{V} = (V_{I_1}, \cdots, V_{I_n})$ 等可能地等于 $v_1, \cdots, v_n$ 的 n ! 个排列中的任意一个, 并令 $\boldsymbol{V_1} = (V_{I_n}, \cdots, V_{I_1})$.

定理 如果 g 和 h 既不都是定义在 $v_1, \cdots, v_n$ 的排列上的可交换减函数也不都是定义在 $v_1, \cdots, v_n$ 的排列上的可交换增函数, 则

$$\mathrm{Cov}(g(\boldsymbol{V}), h(\boldsymbol{V_1})) \leqslant 0$$

我们应用如下的引理证明这个定理.

引理 如果 h 是可交换增函数, 则 $E[h(\boldsymbol{V})|I_1 = i]$ 是 i 的增函数, 并且 $E[h(\boldsymbol{V_1})|I_1 = i]$ 是 i 的减函数. 如果 h 是可交换减函数, 则 $E[h(\boldsymbol{V})|I_1 = i]$ 是 i 的减函数, 并且 $E[h(\boldsymbol{V_1})|I_1 = i]$ 是 i 的增函数.

204

引理的证明 令 $i>1$, $\boldsymbol{P}$ 表示 $1,\cdots,i-2,i+1,\cdots,n$ 的所有 $(n-2)!$ 个排列的集合. 则对可交换增函数 h

$$\begin{aligned}E[h(\boldsymbol{V})|I_1=i]&=\frac{1}{(n-1)!}\sum_{(x_1,\cdots,x_{n-2})\in\boldsymbol{P}}\sum_{k=1}^{n-1}h(v_i,v_{x_1},\cdots,v_{x_{k-1}},v_{i-1},v_{x_k},\cdots,v_{x_{n-2}})\\&\geqslant\frac{1}{(n-1)!}\sum_{(x_1,\cdots,x_{n-2})\in\boldsymbol{P}}\sum_{k=1}^{n-1}h(v_{i-1},v_{x_1},\cdots,v_{x_{k-1}},v_i,v_{x_k},\cdots,v_{x_{n-2}})\\&=E[h(\boldsymbol{V})|I_1=i-1]\end{aligned}$$

其中不等式成立是由于 $v_{i-1}<v_i$ 和 h 为可交换增函数. 类似有

$$\begin{aligned}E[h(\boldsymbol{V}_1)|I_1=i]&=\frac{1}{(n-1)!}\sum_{(x_1,\cdots,x_{n-2})\in\boldsymbol{P}}\sum_{k=1}^{n-1}h(v_{x_1},\cdots,v_{x_{k-1}},v_{i-1},v_{x_k},\cdots,v_{x_{n-2}},v_i)\\&\leqslant\frac{1}{(n-1)!}\sum_{(x_1,\cdots,x_{n-2})\in\boldsymbol{P}}\sum_{k=1}^{n-1}h(v_{x_1},\cdots,v_{x_{k-1}},v_i,v_{x_k},\cdots,v_{x_{n-2}},v_{i-1})\\&=E[h(\boldsymbol{V}_1)|I_1=i-1]\end{aligned}$$

对可交换减函数 h 成立. □

定理的证明 对 n 进行归纳证明. 假设 g 和 h 是可交换增函数. 当 $n=1$ 时, 定理显然成立 (因为在这种情况下, 随机向量是确定的, 协方差是 0). 假设对 $n-1$ 定理成立. 因为, 对固定的 j, $g(v_j,x_1,\cdots,x_{n-1})$ 和 $h(x_1,\cdots,x_{n-1},v_j)$ 都是定义在 $v_1,\cdots,v_{j-1},v_{j+1}\cdots,v_n$ 的 $(n-1)!$ 个排列 $(x_1,\cdots,x_{n-1})$ 上的可交换增函数, 根据归纳假设有

$$\mathrm{Cov}(g(\boldsymbol{V}),h(\boldsymbol{V}_1)|I_1=j)\leqslant 0$$

于是,

$$E[\mathrm{Cov}(g(\boldsymbol{V}),h(\boldsymbol{V}_1)|I_1)]\leqslant 0$$

根据引理 1 知, $E[g(\boldsymbol{V})|I_1]$ 是 I_1 的增函数, 并且 $E[h(\boldsymbol{V}_1)|I_1]$ 是 I_1 的减函数, 于是有

$$\mathrm{Cov}(E[g(\boldsymbol{V}|I_1],E[h(\boldsymbol{V}_1)|I_1])\leqslant 0$$

205

上述结论成立的根据是条件协方差恒等式, 即对任意随机变量 X,Y,Z,

$$\mathrm{Cov}(X,Y)=E[\mathrm{Cov}(X,Y)|Z]+\mathrm{Cov}(E[X|Z],E[Y|Z])$$

当 g 和 h 是可交换减函数时, 证明类似. □

在 $g = h$ 时定理表明，当 f 是可交换单调函数时，用 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}$ 估计 f 比用与 $\boldsymbol{V}$ 独立的其他随机排列和 $\boldsymbol{V}$ 估计 f 要好. 例如，在例 8z 中，当工作被处理的顺序为随机的时，我们想用模拟估计期望回报，则使用"对立的排列"比起产生独立的排列的方法更能达到方差缩减的目的.

注　如果去掉 g 的条件，该定理不成立. 例如，假设当 1 既不在排列的前端也不在排列的后端时，定义在 $1, \cdots, n$ 的排列上的函数 f 的值很大. 则当 $I_1, \cdots, I_n$ 是 $1, \cdots, n$ 的一个随机排列时，$f(I_1, \cdots, I_n)$ 与 $f(I_n, \cdots, I_1)$ 有可能正相关.

随机子集

假设现在我们想模拟估计 $\theta = E[g(B)]$，其中 B 等可能地等于 $S = \{1, 2, \cdots, n\}$ 的含 k 个元素的 $\binom{n}{k}$ 个子集中的任何一个，g 是定义在 S 的含 k 个元素的子集上的函数. 如果对所有由 $k-1$ 个元素组成的 S 的子集 A，当 $i \notin A$ 时，$g(A \cup i)$ 是 i 的增（减）函数，称函数 g 是增（减）函数. 现在，不用通过产生含 k 个元素的独立的子集对 θ 进行估计，而是先随机地产生一个 S 的含 k 个元素的子集，称之为 R_1；然后，随机地产生一个 $S - R_1$ 的含 k 个元素的子集，称之为 R_2；然后，随机地产生一个 $S - R_1 - R_2$ 的含 k 个元素的子集，称之为 R_3，以此类推. 应用上面证明的定理，我们可以证明，当 g 是单调函数时，后面这个方法得到的 θ 估计比用产生独立的子集得到的结果要好.

推论　*在上述的 R_i 给定的情况下，当 g 是增函数或者是减函数时，*

$$\operatorname{Cov}(g(R_i), g(R_j)) \leqslant 0, \ i \neq j$$

证明　假设 $n \geqslant 2k$. 在 S 的排列上定义函数 f 如下：

206

$$f(i_1, \cdots, i_n) = g(i_1, \cdots, i_k)$$

其中 $g(i_1, \cdots, i_k) = g(\{i_1, \cdots, i_k\})$. 因为 f 是可交换的单调函数，从定理 2 可得，对于一个随机排列 $I_1, \cdots, I_n$，$f(I_1, \cdots, I_n)$ 与 $f(I_n, \cdots, I_1)$ 的协方差是非正的，即

$$\operatorname{Cov}(g(I_1, \cdots, I_k), g(I_{n-k+1}, \cdots, I_n)) \leqslant 0$$

结论成立是因为 $g(I_1, \cdots, I_k)$ 与 $g(I_{n-k+1}, \cdots, I_n)$ 的联合分布与 $g(R_i)$ 和 $g(R_j)$ 的联合分布相同，其中 R_i 与 R_j 是从 S 中随机选取的、不相交的、含 k 个元素的子集. □

注 (a)　如果 n 不是 k 的整数倍，即 $n = ki + j$，其中 $0 < j < k$，则产生 $R_1, \cdots, R_i$ 之后，可以结束，但最好是应用剩余的不属于 $R_1, \cdots, R_i$ 中任何一个的 j 个元素，加上从 $\bigcup_{t=1}^{i} R_t$ 中随机选取的 $k-j$ 个元素，产生一个含 k 个元素的 S 的子集.

注 (b) 如果没有 g 的一些条件，则该推论不成立. 例如，假设 $n=2k$，且当 R 恰好包含 1, 2 中的一个时，$g(R)$ 的值大 (比如说等于 1)；否则，$g(R)$ 的值小 (比如说等于 0). 则 $g(R)$ 和 $g(R^C)$ 可能正相关.

8.10 附录：在估计单调函数期望值时，对偶变量法的证明

下面定理是说明当函数 h 是其任一个自变量的单调函数时，使用对偶变量法可以缩减方差（相对于产生新的独立随机数集）的关键.

定理 如果随机变量 $X_1,\cdots,X_n$ 是相互独立的，则对任何有 n 个自变量的单调增函数 f 和 g 有

$$E[f(\boldsymbol{X})g(\boldsymbol{X})]\geqslant E[f(\boldsymbol{X})]E[g(\boldsymbol{X})] \tag{8.19}$$

其中 $\boldsymbol{X}=(X_1,\cdots,X_n)$.

证明 用归纳法证明. 先证 $n=1$ 时的情况. 设 f 和 g 都是单变量的单调增函数，则对任意的 x 和 y 有

$$[f(x)-f(y)][g(x)-g(y)]\geqslant 0$$

如果 $x\geqslant y(x\leqslant y)$，则上述两个因式都是非负（非正）的. 于是，对任何随机变量 X 和 Y，

$$[f(X)-f(Y)][g(X)-g(Y)]\geqslant 0$$

意味着

$$E\{[f(X)-f(Y)][g(X)-g(Y)]\}\geqslant 0$$

或等价于

$$E[f(X)g(X)]+E[f(Y)g(Y)]\geqslant E[f(X)g(Y)]+E[f(Y)g(X)]$$

如果假设随机变量 X 和 Y 又是独立同分布的，则在这种情况下，有

$$E[f(X)g(X)]=E[f(Y)g(Y)]$$
$$E[f(X)g(Y)]=E[f(Y)g(X)]\ =\ E[f(X)]E[g(X)]$$

因此，当 $n=1$ 时结论成立.

假定等式 (8.19) 对 $n-1$ 个变量的情况成立，现在假设随机变量 $X_1,\cdots,X_n$ 是相互独立的，并且 f 和 g 是单调增函数. 则

$$
\begin{aligned}
&E[f(\boldsymbol{X})g(\boldsymbol{X})|X_n = x_n] \\
&= E[f(X_1,\cdots,X_{n-1},x_n)g(X_1,\cdots,X_{n-1},x_n)|X_n = x_n] \\
&= E[f(X_1,\cdots,X_{n-1},x_n)g(X_1,\cdots,X_{n-1},x_n)] \qquad \text{根据独立性} \\
&\geqslant E[f(X_1,\cdots,X_{n-1},x_n)]E[g(X_1,\cdots,X_{n-1},x_n)] \qquad \text{根据递归假设} \\
&= E[f(\boldsymbol{X})|X_n = x_n]E[g(\boldsymbol{X})|X_n = x_n]
\end{aligned}
$$

因此，

$$E[f(\boldsymbol{X})g(\boldsymbol{X})|X_n] \geqslant E[f(\boldsymbol{X})|X_n]E[g(\boldsymbol{X})|X_n]$$

对两端取期望，得

$$
\begin{aligned}
E[f(\boldsymbol{X})g(\boldsymbol{X})] &\geqslant E[E[f(\boldsymbol{X})|X_n]E[g(\boldsymbol{X})|X_n]] \\
&\geqslant E[f(\boldsymbol{X})]E[g(\boldsymbol{X})]
\end{aligned}
$$

208

最后一个不等式成立是因为 $E[f(\boldsymbol{X})|X_n]$ 和 $E[g(\boldsymbol{X})|X_n]$ 都是 X_n 的单调增函数. 根据 $n=1$ 时的结论得

$$
\begin{aligned}
E[E[f(\boldsymbol{X})|X_n]E[g(\boldsymbol{X})|X_n]] &\geqslant E[E[f(\boldsymbol{X})|X_n]]E[E[g(\boldsymbol{X})|X_n]] \\
&= E[f(\boldsymbol{X})]E[g(\boldsymbol{X})]
\end{aligned}
$$

□

推论　如果 $h(x_1,\cdots,x_n)$ 是其每个自变量的单调函数，则对独立随机数 $U_1,\cdots,U_n$，我们有

$$\text{Cov}[h(U_1,\cdots,U_n),h(1-U_1,\cdots,1-U_n)] \leqslant 0$$

证明　不失一般性，通过重新定义 h，可以假定 h 是它的前 r 个自变量的单调增函数，后 $n-r$ 个自变量的单调减函数. 因此，令

$$f(x_1,\cdots,x_n) = h(x_1,\cdots,x_r,1-x_{r+1},\cdots,1-x_n)$$

$$g(x_1,\cdots,x_n) = -h(1-x_1,\cdots,1-x_r,x_{r+1},\cdots,x_n)$$

由此可得 f 和 g 都是单调增函数. 根据上面的定理可知

$$\text{Cov}[f(U_1,\cdots,U_n),g(U_1,\cdots,U_n)] \geqslant 0$$

这等价于

$$
\begin{aligned}
&\text{Cov}[h(U_1,\cdots,U_r,1-U_{r+1},\cdots,1-U_n), \\
&\qquad h(1-U_1,\cdots,1-U_r,U_{r+1},\cdots,U_n)] \leqslant 0.
\end{aligned}
$$

既然随机向量 $h(U_1, \cdots, U_n), h(1-U_1, \cdots, 1-U_n)$ 的联合分布与随机向量

$$h(U_1, \cdots, U_r, 1-U_{r+1}, \cdots, 1-U_n)$$

$$h(1-U_1, \cdots, 1-U_r, U_{r+1}, \cdots, U_n)$$

的联合分布相同, 则可知结论成立. □

习 题

1. 假设我们想要估计 θ, 其中

$$\theta = \int_0^1 e^{x^2} dx$$

证明: 产生随机数 U 并用 $e^{U^2}(1+e^{1-2U})/2$ 估计 θ, 比产生两个随机数 U_1 和 U_2 而用 $(e^{U_1^2}+e^{U_2^2})/2$ 估计 θ 要好.

209

2. 说明如何用对偶变量法得到等式

$$\theta = \int_0^1 \int_0^1 e^{(x+y)^2} dy dx$$

的模拟估计. 在这种情况下, 应用对偶变量模拟估计是否明显地比产生一对随机数更有效?

3. 设 $X_i, i = 1, \cdots, 5$ 是均值为 1 的相互独立的指数随机变量, 考虑如下定义的 θ:

$$\theta = P\left\{\sum_{i=1}^{5} iX_i \geqslant 21.6\right\}$$

(a) 说明如何用应用模拟方法估计 θ.

(b) 给出 θ 的对偶变量估计.

(c) 在这种情况下应用对偶变量法有效吗?

4. 证明: 如果随机变量 X 与 Y 的分布相同, 则 $\text{Var}[(X+Y)/2] \leqslant \text{Var}(X)$, 并据此得出应用对偶变量法无论怎样都不会使方差增加的结论（尽管不要求它与产生独立的随机数集一样有效）.

5. (a) 如果 Z 是标准正态随机变量, 设计一个用对偶变量法估计 $\theta = E[Z^3 e^Z]$ 的方案.

(b) 应用上面的方法, 通过模拟给出一个长度不超过 0.1 的置信度为 95% 的 θ 的置信区间.

6. 设 X 是一个均值为 1 的指数随机变量. 给出一个与 X 负相关的随机变量, 并且它也是均值为 1 的指数随机变量.

7. 证明等式 (8.1).

8. 证明等式 (8.2).

9. 若 $U_n, n \geqslant 1$ 是服从 $(0,1)$ 上均匀分布的独立随机变量序列. 定义

$$S = \min(n : U_1 + \cdots + U_n > 1)$$

可以证明 S 的分布与例 8e 中 N 的分布相同, 因此 $E[S] = e$. 另外, 如果令

210

$$T = \min(n : 1 - U_1 + \cdots + 1 - U_n > 1)$$

则可以证明 $S+T$ 的分布与例 8e 中 $N+M$ 的分布相同. 这建议用 $(S+T+N+M)/4$ 估计 e. 用模拟方法估计 $\text{Var}[(S+T+N+M)/4]$.

10. 在某些情况下, 对均值未知的随机变量 X, 通过模拟可以得到 $P\{X \leqslant a\}$ 的估计 (a 是给定的常数). 一次模拟得到的原始模拟估计是 I, 其中

$$I = \begin{cases} 1 & \text{如果 } X \leqslant a \\ 0 & \text{如果 } X > a \end{cases}$$

因为 I 和 X 显然负相关, 一个自然的想法是用 X 作为控制变量减小方差, 即用 $I + c(X - E[X])$ 形式的估计.

(a) 如果 X 服从 $(0,1)$ 上均匀分布, 确定与原始估计 I 相比方差可能缩减的百分比 (在 c 取最优值时).

(b) 如果 X 服从指数分布, 重做 (a).

(c) 说明为什么 I 和 X 是负相关的.

11. 证明当 α 等于 (8.3) 式所给的值时, $\text{Var}(\alpha X + (1-\alpha)W)$ 的值最小, 并确定这个方差.

12. (a) 说明为什么在习题 1 中, 估计 θ 时可以用控制变量.

(b) 做 100 次模拟运行, 应用 (a) 中的控制变量先估计 c^*, 然后估计所得到的估计量的方差.

(c) 使用与 (b) 中一样的数据, 确定对偶变量估计的方差.

(d) 在本例中, 这两种缩减方差的方法哪种更好?

13. 对习题 2 中给出的 θ 重做习题 12.

14. 对习题 3 中给出的 θ 重做习题 12.

15. 证明在估计 $\theta = E[(1-U^2)^{1/2}]$ 时, 用 U^2 做控制变量比用 U 做控制变量好. 为说明这一点, 用模拟方法估计所需的协方差.

16. 编号为 1, 2, 3, 4, 5 的五个元素最初的排序是随机的 (即开始的排列次序是 1, 2, 3, 4, 5 的一个随机排列). 在每一步, 选取其中一个将其排在前面. 也就是说, 如果当前的排序是 2, 3, 4, 1, 5 并且选取的元素是 1, 则新的排序是 1, 2, 3, 4, 5. 假设各次选取是相互独

211

立的，且在每次选取中，元素 i 被选中的概率是 p_i，其中 $p_1 = 1/15, p_2 = 2/15, p_3 = 3/15, p_4 = 4/15, p_5 = 5/15$. 以 L_j 记第 j 个被选中的元素的位置，并令 $L = \sum_{j=1}^{100} L_j$. 我们感兴趣的是用模拟方法估计 $E[L]$.

(a) 说明如何用模拟方法估计 $E[L]$.

(b) 计算 $E[N_i]$，其中 N_i 是元素 i 在 100 次选取中被选中的次数.

(c) 令 $Y = \sum_{i=1}^{5} iN_i$，你认为 Y 与 L 是正相关还是负相关的?

(d) 提供一个用 Y 做控制变量估计 L 的方案.

(e) 用例 8i 的想法给出一个不同的方法，并提供一个可以确定所给方法的效率的方案.

17. 若 X 和 Y 是相互独立的随机变量，各自的分布分别是 F 和 G，期望分别是 μ_x 和 μ_y. 对给定的值 t，我们感兴趣的是用模拟方法估计 $\theta = P\{X + Y \leqslant t\}$.

(a) 给出原始模拟估计 θ 的方法.

(b) 用条件期望法给出一个改进的估计.

(c) 给出一个可以用来进一步改进 (b) 中估计的控制变量.

18. 假设 Y 是正态随机变量，其均值是 1，方差也是 1. 并假设在 $Y = y$ 的条件下，X 是均值为 y，方差为 4 的正态随机变量. 我们想要应用模拟方法有效地估计 $\theta = P\{X > 1\}$.

(a) 给出原始模拟估计.

(b) 说明如何用条件期望法得到一个改进估计.

(c) 说明如何用对偶变量法对 (b) 的估计做进一步的改进.

(d) 说明如何用控制变量法对 (b) 的估计做进一步的改进.

编一个模拟程序，并用它找到如下估计的方差：

(e) 原始模拟估计.

(f) 条件期望法估计.

(g) 同时使用条件期望法与对偶变量法得到的估计.

(h) 同时使用条件期望法与控制变量法得到的估计.

(i) θ 的确切值是什么?

[提示：独立正态随机变量的和仍是正态随机变量.]

212

19. 在一周时间内，一个意外保险分部被索赔的次数依赖于外部环境因素 U. 如果这个因素的值是 $U = u$，则理赔的次数服从均值是 $\frac{15}{0.5+u}$ 的泊松分布. 假定 U 服从 $(0,1)$ 上的均匀分布，以 p 记在一周时间内被索赔的次数至少是 20 的概率.

(a) 说明如何给出 p 的原始模拟估计.

(b) 给出一个同时使用条件期望法和控制变量法得到的有效估计.

(c) 给出一个同时使用条件期望法和对偶变量法得到的有效估计.

(d) 编一个模拟程序，并用它确定在 (a),(b),(c) 中给出估计的方差.

20. (撞上 - 错过法.) 若 g 是区间 $[0,1]$ 上的有界函数，例如，假设当 $0 \leqslant x \leqslant 1$ 时，$0 \leqslant g(x) \leqslant b$，并且假设我们感兴趣的是用模拟方法求 $\theta = \int_0^1 g(x)\mathrm{d}x$ 的近似值. 要用撞

上－错过方法完成这项任务，需产生一对独立的随机数 U_1 和 U_2. 令 $X = U_1, Y = bU_2$ 以使随机点 (X,Y) 均匀地分布在底是 1，高是 b 的矩形内. 令

$$I = \begin{cases} 1 & \text{如果 } Y < g(x) \\ 0 & \text{其他情况} \end{cases}$$

也就是说，如果随机点 (X,Y) 落在图 8-4 所示的阴影部分内，I 就等于 1.

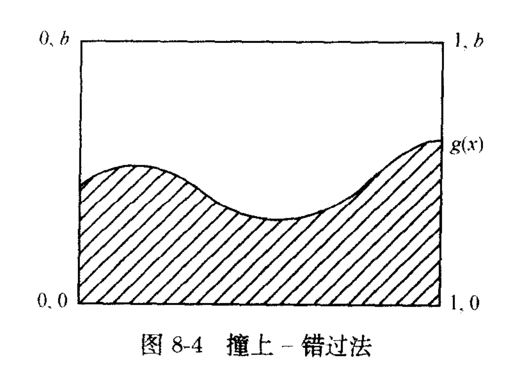


图 8-4　撞上－错过法

(a) 证明 $E[I] = [\int_0^1 g(x)\mathrm{d}x]/b$.

(b) 证明 $\mathrm{Var}(bI) \geqslant \mathrm{Var}(g(U))$，即用撞上－错过法估计的方差比直接用 g 计算的方差大.

213

21. 若 X 和 Y 是相互独立的指数随机变量，X 的期望是 1，Y 的期望是 2，假设我们想用模拟估计 $P\{X+Y>4\}$. 如果想用条件期望法减小估计的方差，则用 X 还是用 Y 做条件？给出理由.

22. 若 X 和 Y 是相互独立的服从二项分布 (n,p) 的随机变量，令 $\theta = E[\mathrm{e}^{XY}]$.

(a) 说明估计 θ 的模拟方法.

(b) 给出一个控制变量，并说明如何用它得到一个比用 (a) 中的原始模拟估计的方差小的估计.

(c) 给出一个直观上看比 (b) 中使用的控制变量更好的控制变量.

[提示：注意应用 $f(x) = \mathrm{e}^x$ 的级数展开.]

(d) 用条件期望法改进原始模拟估计.

(c) 用控制变量法改进 (f) 中的估计.

(f) 用一个有效的方式估计 θ. 到至少有 95% 的把握保证估计的偏差在 0.1 之内时停止模拟.

23. 假设顾客按照参数为 λ 的泊松过程来到一个只有一名服务员的服务站. 服务员空闲时，顾客即刻得到服务，否则要排队等候. 当一个服务完成时，如果队列中还有其他

顾客，则排在第一的顾客接受服务. 所有的顾客的服务时间是独立的随机变量，具有共同的分布 G，假设规定若到时刻 T 的系统是空闲的，服务员就休息一下，否则，就到超过时刻 T 的系统首次变空闲的时刻休息. 用 X 表示从时刻 T 开始到服务员休息的时间间隔，假设我们想用模拟方法估计 $E[X]$. 说明如何利用条件期望法得到 $E[X]$ 的一个有效估计.

[提示：考虑在时刻 T 的模拟，注意在时刻 T 正在接受服务的顾客需要的剩余服务时间和正在排队等待的顾客数 (这个问题需要一些有关 $M/G/1$ 忙期的理论知识).]

24. 考虑顾客以参数为每分钟 2 个的泊松过程来到某单服务员排队系统，顾客接受服务的时间服从均值为 1 分钟的指数分布. 以 T_i 记第 i 个顾客花费在系统中的时间. 我们感兴趣的是用模拟方法估计 $\theta = E[T_1 + \cdots + T_{10}]$.

    (a) 做一个模拟估计原始模拟估计的方差. 也就是估计 $\mathrm{Var}(T_1 + \cdots + T_{10})$.

    (b) 做一个模拟确定通过对偶变量法对原始估计的改进.

    (c) 做一个模拟确定用 $\sum_{i=1}^{10} S_i$ 作为控制变量对原始估计的改进，其中 S_i 是第 i 个服务时间.

    214

    (d) 做一个模拟确定用 $\sum_{i=1}^{10} S_i - \sum_{i=1}^{9} I_i$ 作为控制变量对原始估计的改进，其中 I_i 是第 i 个与第 $i+1$ 个到达者之间的时间间隔.

    (e) 做一个模拟确定用 $\sum_{i=1}^{10} E[T_i|N_i]$ 作为估计对原始估计的改进，其中 N_i 是第 i 个顾客到达时系统中的顾客数 (因此 $N_1 = 0$).

25. 用例 8m 使用的方差缩减技术重做第 5 章习题 10. 如同估计未使用方差缩减技术的估计量的方差一样估计新得到的估计量的方差.

26. 计算例 8r 中的 $E[X|i], i = 0, 1, 2$.

27. 在例 8r 描述的视频纸牌游戏模型中，估计期望收益的原始模拟估计的方差. 然后估计用方差缩减技术得到的估计量的方差. 你给出的期望收益的估计是什么？(如果它小于 1，则这款游戏对玩家来说是不公平的.)

28. 考虑一个由 20 个独立部件组成的系统，每个部件发生故障的概率是 $0.5 + i/50, i = 1, \cdots, 20$. 以 X 记发生故障的部件数. 用模拟方法有效地估计 $P\{X \leqslant 5\}$.

29. 在上题中，估计 $P\{X = 5|X \leqslant 5\}$.

30. 如果 X 满足 $P\{0 \leqslant X \leqslant a\} = 1$，证明：

    (a) $E[X^2] \leqslant aE[X]$.

    (b) $\mathrm{Var}(X) \leqslant E[X](a - E[X])$.

    (c) $\mathrm{Var}(X) \leqslant a^2/4$.

    [提示：注意到 $\max_{0 \leqslant p \leqslant 1} p(1-p) = 1/4$.]

31. 在例 8x 中，给出 $P\{S > 62\}$ 的解析上界.

32. 用模拟方法估计例 8x 中的 $E[S|S>200]$.

33. 假设有一个“暗箱”，对它发指令就会产生参数为 $(3/2,1)$ 的伽玛分布的随机变量. 说明怎样用暗箱求 $E[e^X/(X+1)]$ 的近似值，其中 X 是均值为 1 的指数随机变量.

34. 假设在第 6 章习题 13 中，我们感兴趣的是用模拟方法估计 p，即到某一固定的时刻 t 系统失效的概率. 如果 p 很小，说明如何用重要抽样法得到一个比原始模拟估计更有效的估计. 选择一些使 p 较小的 α, C 和 t 的值，如同估计原始模拟估计的方差一样，做一个模拟估计重要抽样法估计的方差.

215

35. 考虑用两种方法制造同一种产品. 使用不同的方法的利润依赖于参数 α 的值，以 $v_i(\alpha)$ 记使用方法 i 的利润，它是 α 的函数. 假设当 α 的值较小时，方法 1 的效益最好是由于 $v_1(\alpha)$ 是 α 的减函数，但是当 α 的值较大时，方法 2 的效益最好是由于 $v_2(\alpha)$ 是 α 的增函数. 如果平常 α 的值是来自分布为 F 的随机变量，若比较这两种方法的平均利润，我们是产生一个 α 的值并用这个 α 计算利润，还是产生 α_1 和 α_2 然后计算 $v_i(\alpha_i), i=1,2$?

36. 考虑有 n 个名字的一个名单，其中 n 很大，并假设某给定的名字多次出现在名单上. 以 $N(i)$ 记第 i 个位置上的名字出现在名单上的次数，$i=1,\cdots,n$，以 D 记在名单上的不同名字的个数. 我们感兴趣的是用模拟方法估计 $\theta=E[D]$.

    (a) 证明 $\theta=\sum_{i=1}^{n}\frac{1}{N(i)}$.

    令 X 等可能地取 $i=1,\cdots,n$ 中每一个值. 确定在位置 X 上的名字，并且从开头起找这个名字，直至找到为止，如果这个名字第一次出现是在位置 X 上，则令 $Y=1$，否则，令 $Y=0$.

    (b) 证明 $E[Y|N(X)]=\frac{1}{N(X)}$.

    (c) 证明 $E[nY]=\theta$.

    (d) 现在，如果位置 X 是这个名字在名单上最后一次出现的位置，则令 $W=1$，否则，令 $W=0$. (也就是说，如果在名单上从后往前找，这个名字首次出现是在位置 X 上，则 $W=1$.) 证明 $n(W+Y)/2$ 是 θ 的一个无偏估计.

    (e) 证明，如果每个名字在名单上至少出现两次，则 (d) 中的估计比估计 $n(Y_1+Y_2)/2$ 好，其中 Y_1 和 Y_2 相互独立且与 Y 有相同的分布.

    (f) 证明 $n/(N(X))$ 的方差比 (e) 中的估计的方差小，尽管当多次重复时，(e) 中的估计仍可能是更有效的，因为它的搜索速度更快.

37. 以 $\Phi^{-1}(x)$ 记标准正态分布函数 $\Phi(x)$ 的反函数. 假定我们可以有效地计算函数 $\Phi(x)$ 和 $\Phi^{-1}(x)$，证明可以通过产生随机数 U，令 $Y=U+(1-U)\Phi(c)$，并且令

$$Z=\Phi^{-1}(Y)$$

产生以大于 c 为条件的标准正态随机变量 Z. [Excel 有内嵌的计算 $\Phi(x)$ 和 $\Phi^{-1}(x)$ 的程序.]

216

38. 就 8.5 节的复合随机向量的估计 $\mathcal{E}$, 证明

$$\mathrm{Var}(\mathcal{E}) \leqslant \mathrm{Var}(g_N(X_1, \ldots, X_N))$$

[提示: 证明 $\mathcal{E}$ 是条件期望估计.]

39. 假设我们想模拟确定 $\theta = E[h(Z_1, \ldots, Z_n)]$, 其中 $Z_1, \ldots, Z_n$ 是相互独立的标准正态随机变量, h 是其每一个分量的增函数. 令 $W = \sum_{i=1}^n a_i Z_i$, 其中 a_i 均非负. 应用下面的引理, 说明我们如何利用基于对 W 分层的分层抽样法, 求 θ 的近似值 (假设可以用逆变换方法模拟 W).

    引理 如果标准正态随机变量 Z 与 X 相互独立, 其中 X 的期望是 μ, 方差是 σ^2, 则在给定 $Z + X = t$ 的条件下, Z 的条件分布仍是正态, 且其期望为 $\frac{t-\mu}{1+\sigma^2}$, 方差为 $\frac{\sigma^2}{1+\sigma^2}$.

40. 当 $h(x_1, \ldots, x_n)$ 是其一些分量的增函数且是其余分量的减函数时, 解释如何使用上一习题的方法.

参考文献

[1] Hammersley, J. M., and D. C. Handscomb, *Monte Carlo Methods.* Wiley, New York, 1964.

[2] Hammersley, J. M., and K. W. Morton, *A New Monte Carlo Technique: Antithetic Variables, Proc. Cambridge Phil. Soc.*, **52,** 449-474 ,1956.

[3] Lavenberg, S. S., and P. D. Welch, *A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations, Management Sci.*, **27,** 322-335, 1981.

[4] Morgan, B. J. T., *Elements of Simulation.* Chapman and Hall, London,1983.

[5] Ripley, B., *Stochastic Simulation.* Wiley, New York, 1986.

[6] Ross, S. M., and K. Lin, *Applying Variance Reduction Ideas in Queuing Simulations, Probability Eng. Informational Sci.*, **15,** 481-494, 2001.

[7] Rubenstein, R. Y., *Simulation and the Monte Carlo Method.* Wiley, New York, 1981.

[8] Siegmund, D., *Importance Sampling in the Monte Carlo Study of Sequential Tests, Ann. Statistics*, **4,** 673-684, 1976.

第 9 章　统计验证技术

引　　言

本章我们将介绍几种验证模拟模型的统计方法. 9.1 节和 9.2 节讨论了拟合优度检验，此检验可用于确定假设的概率分布是否与给定的数据一致. 在 9.1 节中，假设分布完全确定，而在 9.2 节中只确定了其中某些参数，例如，一个均值未知的泊松分布. 9.3 节我们给出如何检验两个样本来自同一总体的假设，即当模拟的数学模型可以准确反映现实问题时，实际数据和模拟数据即为这两个样本. 9.3 节的结果特别有助于检验模拟模型的有效性. 对多个样本情况的推广也在 9.3 节中给出. 最后, 9.4 节给出如何利用真实数据检验产生数据的过程为一个非齐次泊松过程. 对齐次泊松过程的检验也在 9.4 节中给出.

9.1　拟合优度检验

对给定现象作概率分析，是假定这一现象的随机元素中有一些服从特定的概率分布. 例如，对交通网的分析，是假定日常发生事故的数量服从一个泊松分布. 这些假设可以通过观测数据来检验，进而验证假设的概率分布是否与数据一致. 这种检验称为*拟合优度检验*.

219

一种拟合优度检验方法是首先将一个随机变量的可能取值划分为有限个区域. 观测得到变量的一组样本值，当这些数据服从一个特定的概率分布时，将样本中落在每个区域中的个数与理论期望个数进行比较.

本节主要考虑假设分布所有参数已知情况下的拟合优度检验；下一节考虑的是某些参数未知的情况. 首先我们给出一个离散分布的例子，然后是分析连续分布的情况.

1. 离散数据的 χ^2 拟合优度检验

假设观测 n 个独立随机变量, $Y_1, \cdots, Y_n$, 每一个随机变量可能取值为 $1, 2, \cdots, k$. 我们要检验假设 $\{p_i, i = 1, \cdots, k\}$ 为这些随机变量的概率列. 如果用 Y 表示任意的 Y_j，则要检验的假设（也就是原假设）记为 H_0，表示为

$$H_0 : P\{Y = i\} = p_i, \quad i = 1, \cdots, k$$

为检验上述假设，令 N_i，$i = 1, \cdots, k$ 为 Y_j 等于 i 的个数. 由于每个 Y_j 独立

且以 $P\{Y=i\}$ 的概率等于 i，于是，在原假设 H_0 的假设下，N_i 服从参数为 n 和 p_i 的二项分布. 因此，当原假设 H_0 成立时，有

$$E[N_i] = np_i$$

于是可用 $(N_i - np_i)^2$ 来反映 p_i 为 $Y=i$ 的概率的可能性. 当 $(N_i - np_i)^2$ 的值相对于 np_i 很大时，则表明原假设 H_0 不成立. 事实上，由上述原因，可以考虑检验

$$T = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

当 T 很大时，拒绝原假设.

检验量 T 较小时，原假设 H_0 可能成立；而 T 较大时，则意味着原假设 H_0 可能不成立. 现假设实际数据使得 T 取值为 t. 当原假设成立时，考虑 T 取值的可能性，定义为p 值，

$$p\ 值 = P_{H_0}\{T \geqslant t\}$$

220

其中 P_{H_0} 表示在原假设 H_0 成立的前提下求得的概率. 因此，p 值是当原假设成立时 T 大于等于观测所得 t 值的概率. 当 p 值很小时（小于 0.05 的值，或更保守地，小于 0.01 的值通常可作为临界值），即假设与数据值不一致，通常拒绝原假设. 反过来，当 p 值很大时，假设与数据值一致，则接受原假设.

观测检验量的值后，令其为 t，计算

$$p\ 值 = P_{H_0}\{T \geqslant t\}$$

此概率的一个很好的近似可以由一个经典结论获得，当 n 很大时，且原假设 H_0 成立，则 T 近似服从一个自由度为 $k-1$ 的 χ^2 分布. 因此，

$$p\ 值 \approx P\{\chi^2_{k-1} \geqslant t\} \tag{9.1}$$

其中 χ^2_{k-1} 是自由度为 $k-1$ 的 χ^2 随机变量.

例 9a 考虑随机变量，其可能取值为 1,2,3,4,5，我们检验假设随机变量等可能地取得这些值. 即，检验

$$H_0 : p_i = 0.2, \quad i = 1, \cdots, 5$$

如果样本大小为 50 且 N_i 为：

$$12, 5, 19, 7, 7$$

则 p 值的近似值可如下获得. 检验统计量 T 的值为

$$T = \frac{4+25+81+9+9}{10} = 12.8$$

因此,

$$p\text{ 值} \approx P\{\chi_4^2 > 12.8\} = 0.0122$$

221 由于 p 值很小, 所有结果等可能出现的假设被拒绝. □

如果 (9.1) 式给出的 p 值不太小, 等于或大于 0.15 , 则不能拒绝原假设, 即没有必要再寻找更好的近似. 然而, 当 p 值接近临界值 (如 0.05 或 0.01) 时, 可能就需要一个比 χ^2 近似分布给出的估计值更精确的值. 幸运的是, 更精确的估计可以通过模拟研究获得.

估计 $T = t$ 的 p 值, 其模拟方法如下. 当原假设 H_0 成立时, 为求得 T 大于等于 t 的概率, 我们产生 n 个独立的随机变量 $Y_1^{(1)}, \cdots, Y_n^{(1)}$, 每一个随机变量有概率分布函数 $\{p_i, i = 1, \cdots, k\}$, 即

$$P\{Y_j^{(1)} = i\} = p_i, \quad i = 1, \cdots, k, \qquad j = 1, \cdots, n$$

设

$$N_i^{(1)} = \#\{j : Y_j^{(1)} = i\}$$

且

$$T^{(1)} = \sum_{i=1}^{k} \frac{(N_i^{(1)} - np_i)^2}{np_i}$$

重复这一过程, 再生成 n 个独立的随机变量 $Y_1^{(2)}, \cdots, Y_n^{(2)}$, 每一个随机变量有概率分布函数 $\{p_i, i = 1, \cdots, k\}$, 将其看作第二组数据, 与第一组数据作相同处理, 得出 $T^{(2)}$. 重复这一过程 r 次, 则得到 r 个独立随机变量 $T^{(1)}, T^{(2)}, \cdots, T^{(r)}$, 当原假设 H_0 成立时, 它们与检验统计量 T 有相同的分布. 因此, 由大数定律, 当原假设 H_0 成立时, T_i 大于等于 t 的比例与 T 大于等于 t 的概率近似相等, 即

$$\frac{\#\{l : T^{(l)} \geqslant t\}}{r} \approx P_{H_0}\{T \geqslant t\}$$

2. 连续数据的科尔莫戈罗夫 – 斯米尔诺夫检验

考虑 $Y_1, \cdots, Y_n$ 是独立随机变量的情况, 要检验的原假设为这些随机变量有相同的分布函数 F, 其中 F 是一个给定的连续分布函数. 一种检验 H_0 的方法是将 Y_j 的可能取值所构成的集合划分为 k 个不同的区间, 即

222

$$(y_0, y_1), (y_1, y_2), \cdots, (y_{k-1}, y_k), \text{其中 } y_0 = -\infty,\ y_k = +\infty$$

然后考虑离散化随机变量 Y_j^d, $j=1,\cdots,n$, 定义为

$$Y_j^d = i, \quad \text{如果 } Y_j \text{ 落在区间 } (y_{i-1}, y_i)\text{上}$$

则原假设为

$$P\{Y_j^d = i\} = F(y_i) - F(y_{i-1}), \quad i = 1,\cdots,k$$

且可使用前面介绍的χ^2 拟合优度检验.

还有另一种比离散化更有效的方法检验 Y_j 来自连续分布函数 F, 其具体步骤如下：观测 $Y_1,\cdots,Y_n$, 设 F_e 为经验分布函数, 定义如下

$$F_e(x) = \frac{\#\{i : Y_i \leqslant x\}}{n}$$

即, $F_e(x)$ 为观测值小于或等于 x 的比例. 由于 $F_e(x)$ 是观测值小于或等于 x 的概率的一个自然估计, 因此, 当原假设 F 为真实分布成立时, $F_e(x)$ 应该接近 F. 由于这一结论对于所有 x 都成立, 原假设 H_0 的一个自然检验量为

$$D \equiv \max_x |F_e(x) - F(x)|$$

其中 x 的取值区间为 $(-\infty, +\infty)$. 称 D 为科尔莫戈罗夫 – 斯米尔诺夫检验统计量.

已知数据 $Y_j = y_j, j=1,\cdots,n$, 计算 D 的值, 定义 $y_{(1)}, y_{(2)},\cdots,y_{(n)}$ 为 y_j 的递增排序. 即,

$$y_{(j)} = y_1,\cdots,y_n\text{中的第 } j \text{ 个最小值}$$

例如, 如果 $n=3$ 且 $y_1=3$, $y_2=5$, $y_3=1$, 则 $y_{(1)}=1, y_{(2)}=3, y_{(3)}=5$. 由于 $F_e(x)$ 可写成

$$F_e(x) = \begin{cases} 0 & x < y_{(1)} \\ \dfrac{1}{n} & y_{(1)} \leqslant x < y_{(2)} \\ \vdots & \\ \dfrac{j}{n} & y_{(j)} \leqslant x < y_{(j+1)} \\ \vdots & \\ 1 & y_{(n)} \leqslant x \end{cases}$$

223

在区间 $(y_{(j-1)}, y_{(j)})$ 中 $F_e(x)$ 为常数, 在 $y_{(1)},\cdots,y_{(n)}$ 点处跳跃 $1/n$. 由于 $F(x)$ 是以 1 为界的关于 x 的递增函数, 因此 $F_e(x)-F(x)$ 的最大值非负, 并在

$y_{(j)}, j=1,\cdots,n$ 点处取得，见图 9-1. 即，

$$\max_x\{F_e(x)-F(x)\}=\max_{j=1,\cdots,n}\left\{\frac{j}{n}-F(y_{(j)})\right\} \tag{9.2}$$

同样，$F(x)-F_e(x)$ 的最大值也是非负的，并在每个跳跃点 $y_{(j)}$ 之前取得，于是，

$$\max_x\{F(x)-F_e(x)\}=\max_{j=1,\cdots,n}\left\{F(y_{(j)})-\frac{(j-1)}{n}\right\} \tag{9.3}$$

由 (9.2) 和 (9.3) 式，有

$$\begin{aligned} D&=\max_x|F_e(x)-F(x)| \\ &=\max\{\max\{F_e(x)-F(x)\},\max\{F(x)-F_e(x)\}\} \\ &=\max\left\{\frac{j}{n}-F(y_{(j)}),F(y_{(j)})-\frac{(j-1)}{n},j=1,\cdots,n\right\} \end{aligned} \tag{9.4}$$

(9.4) 式可用于计算 D 值.

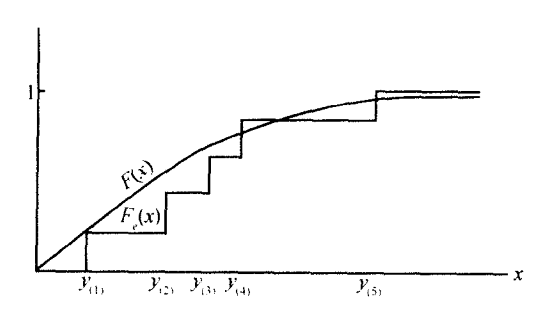


图 9-1　$n=5$ 时的经验分布函数

现假设由观测值 Y_j 计算得 $D=d$. 由于当 D 值很大时，与 F 为真实分布的原假设不一致，因此这组数据的 p 值为

224

$$p\text{ 值}=P_F\{D\geqslant d\}$$

其中 P_F 表示这一概率是在假设 H_0 成立的前提下计算的（即 F 是真实分布）.

上述 p 值可根据下面命题模拟近似. 该命题表明 $P_F\{D\geqslant d\}$ 不依赖于真实分布 F. 利用这一结论，我们可选择任一连续分布 F 做模拟来估计 p 值 [于是我们可选择（0, 1）均匀分布].

命题　对任一连续分布 F，$P_F\{D\geqslant d\}$ 是相同的.

证明

$$\begin{aligned}P_F\{D \geqslant d\} &= P_F\left\{\max_x\left|\frac{\#\{i:Y_i \leqslant x\}}{n} - F(x)\right| \geqslant d\right\}\\&= P_F\left\{\max_x\left|\frac{\#\{i:F(Y_i) \leqslant F(x)\}}{n} - F(x)\right| \geqslant d\right\}\\&= P\left\{\max_x\left|\frac{\#\{i:U_i \leqslant F(x)\}}{n} - F(x)\right| \geqslant d\right\}\end{aligned}$$

其中 $U_1,\cdots,U_n$ 为来自均匀分布 (0, 1) 的独立随机变量. 由于 F 是递增函数, 因此 $Y \leqslant x$ 等价于 $F(Y) \leqslant F(x)$, 第一个等式成立; 又如果 Y 有连续分布 F, 则随机变量 $F(Y)$ 服从 (0, 1) 均匀分布 (证明留作练习), 因此, 可得第二个等式.

接下来, 我们令 $y = F(x)$, 并注意到 x 取值范围从 $-\infty$ 到 $+\infty$, $F(x)$ 取值范围从 0 到 1, 则有

$$P_F\{D \geqslant d\} = P\left\{\max_{0\leqslant y\leqslant 1}\left|\frac{\#\{i:U_i \leqslant y\}}{n} - y\right| \geqslant d\right\}$$

此式表明, 当原假设 H_0 成立时, D 的分布不依赖于实际分布 F. □

由前面命题可以看出, 在由数据确定 D 值后, 令 $D = d$, p 值可通过模拟 (0,1) 均匀分布随机变量获得. 即, 我们产生 n 个随机数 $U_1,\cdots,U_n$ 并检查不等式

$$\max_{0\leqslant y\leqslant 1}\left|\frac{\#\{i:U_i \leqslant y\}}{n} - y\right| \geqslant d \tag{9.5}$$

是否成立. 多次重复这一过程, 不等式成立次数的比例即为此组数据 p 值的估计. [225] 正如前面提到的, 不等式 (9.5) 的左边可由排序后的随机数计算得到, 然后利用等式

$$\max\left|\frac{\#\{i:U_i \leqslant y\}}{n} - y\right| = \max\left\{\frac{j}{n} - U_{(j)}, U_{(j)} - \frac{(j-1)}{n}, j = 1,\cdots,n\right\}$$

其中 $U_{(j)}$ 为 $U_1,\cdots,U_n$ 中的第 j 个最小值. 例如, 如果 $n = 3$ 且 $U_1 = 0.7$, $U_2 = 0.6$, $U_3 = 0.4$, 则 $U_{(1)} = 0.4$, $U_{(2)} = 0.6$, $U_{(3)} = 0.7$ 且此组数据的 D 值为:

$$D = \max\left\{\frac{1}{3} - 0.4, \frac{2}{3} - 0.6, 1 - 0.7, 0.4, 0.6 - \frac{1}{3}, 0.7 - \frac{2}{3}\right\} = 0.4$$

例 9b 检验假设总体服从均值为 100 的指数分布, 即 $F(x) = 1 - e^{-x/100}$. 如果来自此分布的 10 个样本值 (排序后) 为

66, 72, 81, 94, 112, 116, 124, 140, 145, 155

则可得到什么结论?

为回答上述问题, 我们首先利用 (9.4) 式计算科尔莫戈罗夫 - 斯米尔诺夫检验量 D 的值. 计算后可得 $D = 0.483\,148\,7$. 为获得 p 值的近似值, 我们进行一次模拟并得到如下输出结果:

```
RUN
THIS PROGRAM USES SIMULATION TO APPROXIMATE THE p-value
   OF THE KOLMOGOROV-SMIRNOV TEST
Random number seed (-32768 to 32767)?  4567
ENTER THE VALUE OF THE TEST QUANTITY
?0.4831487
ENTER THE SAMPLE SIZE
?  10
ENTER THE DESIRED NUMBER OF SIMULATION RUNS
?  500
THE APPROXIMATE p-value IS 0.012
OK
```

由于 p 值很小 (来自均值为 100 的指数分布的 10 个值, 其最小为 66 的可能性很小), 故拒绝原假设. □

226

9.2 参数未知情况下的拟合优度检验

1. 离散数据情况

对概率 $\{p_i, i = 1, \cdots, k\}$ 不完全已知的原假设也可使用拟合优度检验. 例如, 要检验某一地区日常交通事故次数是否服从均值未知的泊松分布. 为检验这一假设, 考虑 n 天的数据并令 Y_i 为第 i 天的交通事故次数, $i = 1, \cdots, n$. 要确定这些数据是否与泊松分布的假设一致, 其困难首先在于如果泊松分布的假设成立, 则这些数据会有无穷多的可能取值. 然而, 通过将这些可能取值划分到有限的 k 个区域中, 并且观测这 n 个数据位于哪个区域, 这一问题就可以解决. 例如, 如果要检验的区域很小, 且日常交通事故不多, 则我们可以如下划分区域: 如某天发生交通事故 $i-1$ 次, 则称此天发生的事故次数落入区域 i, $i = 1, 2, 3, 4, 5$, 而区域 6 则表示发生交通事故 5 次或 5 次以上. 因此, 如果真实分布是均值为 λ 的泊松分布, 则有

$$p_i = P\{Y = i - 1\} = \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!}, \quad i = 1, 2, 3, 4, 5 \tag{9.6}$$

$$p_6 = 1 - \sum_{j=0}^{4} \frac{e^{-\lambda}\lambda^j}{j!}$$

对真实分布为泊松分布的原假设，获得其拟合优度检验的另一个困难在于均值 λ 未知. 当 λ 未知时，直觉上要从数据中估计 λ，(记为 $\hat{\lambda}$)，并计算检验统计量的值

$$T = \sum_{i=1}^{k} \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

其中 N_i 为落入区域 i 中的 Y_j 的次数，$\hat{p}_i$ 为在原假设 H_0 下，Y_j 落入区域 $i(i = 1,\cdots,k,)$ 的概率估计，此估计可通过将表达式 (9.6) 中的 λ 替换为 $\hat{\lambda}$ 得到.

上述方法适用于原假设中有未知参数且需要计算 $p_i(i = 1,\cdots,k)$ 的情况. 现假设有 m 个未知参数. 可以证明，如果对这些参数的估计合理，在原假设成立且 n 很大时，检验量 T 服从自由度为 $k-1-m$ 的 χ^2 分布. （换言之，每一个被估参数损耗一个自由度.）

227

如果检验量取值 $T = t$，则使用上述方法，近似计算 p 值为

$$p\text{ 值} \approx P\{\chi^2_{k-1-m} \geqslant t\}$$

其中 χ^2_{k-1-m} 是自由度为 $k-1-m$ 的 χ^2 随机变量.

例 9c 假设在 30 天内有 6 天没有发生交通事故，2 天发生 1 次交通事故，1 天发生 2 次交通事故，9 天发生 3 次交通事故，7 天发生 4 次交通事故，4 天发生 5 次交通事故，1 天发生 8 次交通事故. 要检验数据是否与泊松分布的原假设一致，首先注意到由于总共发生 87 次交通事故，因此泊松分布的均值估计为

$$\hat{\lambda} = \frac{87}{30} = 2.9$$

于是 $P\{Y = i\}$ 的估计为 $e^{-2.9}(2.9)^i/i!$，我们因此获得在本节开始提到的有 6 个区域的概率估计的值

$$\hat{p}_1 = 0.0500, \quad \hat{p}_2 = 0.1596 \quad \hat{p}_3 = 0.2312$$

$$\hat{p}_4 = 0.2237, \quad \hat{p}_5 = 0.1622 \quad \hat{p}_6 = 0.1682$$

使用数据 $N_1 = 6$，$N_2 = 2$，$N_3 = 1$，$N_4 = 9$，$N_5 = 7$，$N_6 = 5$，我们得到统计量的值

$$T = \sum_{i=1}^{6} \frac{(N_i - 30\hat{p}_i)^2}{30\hat{p}_i} = 19.887$$

为确定 p 值，运行 9.1 节中的程序，得到

$$p \text{ 值} \approx P\{\chi_4^2 > 19.887\} = 0.0005$$

因此泊松分布的原假设被拒绝. □

我们同样也可以用模拟估计 p 值. 然而，由于原假设不再是完全已知的概率模型，利用模拟来决定检验量的 p 值就比原先要复杂. 其具体做法如下.

228

(a) **模型**. 令原假设是数据 $Y_1, \cdots, Y_n$ 组成的样本且来自含有未知参数 $\theta_1, \cdots, \theta_m$ 的分布. 同时，当原假设成立时，令 Y_i 可能的取值为 $1, \cdots, k$.

(b) **初始步骤**. 使用样本数据估计未知参数. 特别地，令 $\hat{\theta}_j$ 为 θ_j 的估计，$j = 1, \cdots, m$. 现计算检验统计量的值

$$T = \sum_{i=1}^{k} \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

其中 N_i 是等于 i 的数据个数，$i = 1, \cdots, k$，且当 $\hat{\theta}_j$ 代替 θ_j 时，$\hat{p}_i$ 为 p_i 的估计，$j = 1, \cdots, m$. 令 t 为检验统计量 T 的值.

(c) **模拟步骤**. 现做一系列模拟来估计这组数据的 p 值. 首先注意到所有模拟使用的是原假设为真时的总体分布并且 θ_j 与步骤 (b) 中得到的估计 $\hat{\theta}_j$ 相等，$j = 1, \cdots, m$.

模拟一个来自上述总体分布，大小为 n 的样本，并令 $\hat{\theta}_j(\text{sim})$ 为基于模拟数据得到的关于 θ_j 的估计，$j = 1, \cdots, m$. 现确定值

$$T_{\text{sim}} = \sum_{i=1}^{k} \frac{[N_i - n\hat{p}_i(\text{sim})]^2}{n\hat{p}_i(\text{sim})}$$

其中 N_i 是等于 i 的模拟数据个数，$i = 1, \cdots, k$，且当 $\hat{\theta}_j(\text{sim})$ 与 θ_j 相等时，$\hat{p}_i(\text{sim})$ 等于 p_i，$j = 1, \cdots, m$.

模拟步骤要多次重复. 则 p 值的估计为 T_{sim} 大于等于 t 的比例.

例 9d　再次考虑例 9c. 由例中的数据得到估计 $\hat{\lambda} = 2.9$ 及检验量 $T = 19.887$. 现模拟生成均值为 2.9 的 30 个独立泊松随机变量并计算

$$T^* \equiv \sum_{i=1}^{6} \frac{(X_i - 30p_i^*)^2}{30p_i^*}$$

其中 X_i 为 30 个数中落入区域 i 的个数，p_i^* 为一个泊松随机变量落入区域 i 的概率，这一随机变量的均值取 30 个生成数的平均值. 多次重复模拟步骤，p 值的估计为 T^* 大于等于 19.887 次数的比例. □

229

2. 连续数据情况

现考虑检验原假设为随机变量 $Y_1, \cdots, Y_n$ 来自连续分布函数 $F_{\boldsymbol{\theta}}$，其中 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)$ 为未知参数向量. 例如，要检验 Y_j 来自一个正态分布总体. 要利用科尔莫戈罗夫－斯米尔诺夫检验，首先要利用数据估计参数向量 $\boldsymbol{\theta}$，记向量估计为 $\hat{\boldsymbol{\theta}}$. 检验统计量 D 的值计算如下

$$D = \max_x |F_e(x) - F_{\hat{\boldsymbol{\theta}}(x)}|$$

其中 $\hat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 的估计，分布函数 $F_{\hat{\boldsymbol{\theta}}}$ 可由 $F_{\boldsymbol{\theta}}$ 获得.

如果统计量的值 $D = d$，p 值可由 $P_{F_{\hat{\theta}}}\{D \geqslant d\} = P_U\{D \geqslant d\}$ 粗略近似. 即，确定 D 值后，可获得 p 值粗略的近似，而实际上高估了 p 值. 如果得到的 p 值估计不是很小，则不拒绝原假设，检验停止. 然而，如果 p 值的估计很小，则有必要使用模拟方法对 p 值的真值做更精确的估计. 其具体步骤如下.

1. 利用数据估计 $\boldsymbol{\theta}$，设为 $\hat{\boldsymbol{\theta}}$. 计算前面所讲的 D 值.

2. 使用分布 $F_{\hat{\boldsymbol{\theta}}}$ 做模拟. 生成样本大小为 n 的分布并设 $\hat{\boldsymbol{\theta}}(\text{sim})$ 为模拟得到的 $\boldsymbol{\theta}$ 的估计. 计算

$$\max_x |F_{e,\text{sim}}(x) - F_{\hat{\boldsymbol{\theta}}(\text{sim})}(x)|$$

其中 $F_{e,\text{sim}}$ 为模拟数据的经验分布函数; 注意这个值是否大于等于 d. 重复这一过程，利用统计量大于等于 d 的次数的比例来估计 p 值.

9.3 两样本问题

假设我们考虑一服务系统的数学模型，要求系统在一天结束时，清除所有顾客；而且，由于连续天的情况是独立同分布的，因此每天的概率分布相同. 模型的一些个别假设，例如，服务次数独立且有相同分布 G，或是顾客到达时间为泊松过程，可以由 9.1 节和 9.2 节的结果分别进行检验. 假设这些检验没有一个 p 值特别小，即分别考虑模型的所有部分，都与系统的真实数据相一致.[我们必须对所谓很小的 p 值格外当心，因为即使模型是正确的，且我们实施了大量的检验，仍有可能得到较小的 p 值. 例如，对独立数据进行 r 次检验，至少一次 p 值小于等于 r 的概率为 $1-(1-\alpha)^r$，即使对很小的 α，当 r 增大时，这一概率仍会变得很大.] 230

然而，到目前为止，我们还不能证明所建立的模型是正确的，且对实际数据有效；整个模型可能是不准确的，因为它不仅包括各个模型部分，而且还包括对各部分之间相互关系的假设. 完整地检验模型的一个方法是将某个全模型的复杂函数作随机变量来考虑. 例如，我们考虑一天之内所有进入系统顾客等待时间的总和. 假设观测系统 m 天并设 $Y_i, i = 1, \cdots, m$，为第 i 天内顾客等待时间的总和. 如果我们

模拟这一模型 n 次，设 $X_i, i=1,\cdots,n$，为在 (模拟的) 第 i 天内，到达系统的所有顾客所用时间的总和. 因为模型假设每天的情况都是独立同分布的，所以随机变量 $X_1,\cdots,X_m$ 有相同的分布，定义为 F. 如果模型准确反映真实系统情况，数据 $Y_1,\cdots,Y_m$ 也应有分布 F. 即模型如果是正确的，模拟数据与真实数据是不可分辨的. 于是，检验完整模型准确性的一个方法是验证 $X_1,\cdots,X_n,Y_1,\cdots,Y_m$ 是独立同分布随机变量，将其作为原假设 H_0. 下面介绍如何检验这一假设.

假设有两组数据 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$，原假设 H_0 为这 $n+m$ 个随机变量独立同分布. 这种统计检验假设问题称为两样本问题.

为检验 H_0，将 $n+m$ 个值 $X_1,\cdots,X_n,Y_1,\cdots,Y_m$ 排序并假设 $n+m$ 个值全不相同，因此排序唯一. 对 $i=1,\cdots,n$，令 R_i 为 $n+m$ 个数据中 X_i 的秩，即，如果 X_i 为 $n+m$ 个值中第 j 个最小的值，则 $R_i=j$. 记统计量

$$R=\sum_{i=1}^{n} R_i$$

231 为第一组数据的秩和. (两组数据中的任何一个都可看作是"第一组"数据.)

如果 R 很大 (说明第一组数据大于第二组数据) 或是很小 (说明第一组数据小于第二组数据)，则都是拒绝原假设的有力证据. 特别地，当 $R=r$ 时，如果

$$P_{H_0}\{R \leqslant r\} \quad 或 \quad P_{H_0}\{R \geqslant r\}$$

很小，我们拒绝原假设. 事实上，在 $R=r$ 时，检验数据的 p 值为

$$p\,值 = 2\min(P_{H_0}\{R \leqslant r\}, P_{H_0}\{R \geqslant r\}). \tag{9.7}$$

[由于当 R 太大或太小时，我们拒绝原假设，于是我们取概率中最小值的两倍作为 p 值. 例如，设在原假设 H_0 下，小于 (大于) 等于 r_* (或 r^*) 的概率为 0.05. 由于在原假设 H_0 下，两事件其中之一发生的概率为 0.1，因此如果 R 值为 r_* (或 r^*)，则 p 值为 0.1.]

通过上述 p 值进行假设检验，即，当 p 值充分小时，拒绝原假设，此检验称为两样本秩和检验. (有时又称为 Wilcoxon 两样本检验或 Mann-Whitney 两样本检验.)

例 9e　设对一系统观测 5 天，得到某量的连续值

342, 448, 504, 361, 453

同时根据系统建立数学模型进行 10 天模拟产生值如下：

186, 220, 225, 456, 276, 199, 371, 426, 242, 311

由于第一组数据的 5 个数据的秩为 8, 12, 15, 9, 13, 所以统计量 $R=57$. □

当 n 和 m 不太大且数据不相同时, 由 (9.7) 式计算 p 值. 如令

$$P_{n,m}(r)=P_{H_0}\{R\leqslant r\}$$

则 $P_{n,m}(r)$ 为第一组数据秩和小于或等于 r 的概率, 其中两组数据同分布, 大小分别为 n 和 m. 根据最大值在第一组数据还是第二组数据中, 我们可以获得求这些概率的递推公式. 如果最大值包含在第一组数据中, 这组数据的秩和为 $n+m$ (最大值的秩) 加上这组数据其他 $n-1$ 个数据的秩和, 同时其他 m 个值来自另一个集合. 所以, 当最大值包含在第一组数据中时, 如果其他 $n-1$ 个数据的秩和小于或等于 $r-n-m$, 则这组数据的秩和小于或等于 r, 并且以概率 $P_{n-1,m}(r-n-m)$ 成立. 如果最大值包含在第二组数据中, 我们有相似的讨论, 第一组数据秩和大于等于 r 的概率为 $P_{n,m-1}(r)$. 最后, 由于 $n+m$ 值中每一个都有可能成为最大值, 所以其包含在第一组数据中的概率为 $n/(n+m)$. 综上所述, 得到递推公式:

232

$$P_{n,m}(r)=\frac{n}{n+m}P_{n-1,m}(r-n-m)+\frac{m}{n+m}P_{n,m-1}(r). \tag{9.8}$$

利用初始边界条件

$$P_{1,0}(k)=\begin{cases}0 & k\leqslant 0\\ 1 & k>0\end{cases} \quad 及 \quad P_{0,1}(k)=\begin{cases}0 & k<0\\ 1 & k\geqslant 0\end{cases}$$

可由公式 (9.8) 递推得到 $P_{n,m}(r)=P_{H_0}\{R\leqslant r\}$ 及 $P_{n,m}(r-1)=1-P_{H_0}\{R\geqslant r\}$.

例 9f 对一系统观测 5 天, 得到某量的如下数据:

132, 104, 162, 171, 129

同时根据系统建立的模型进行 10 天模拟产生值如下:

107, 94, 136, 99, 114, 122, 108, 130, 106, 88

设模型要求这些天的值独立同分布. 要从上述数据确定 p 值, 首先第一组样本的秩和为 R, 即

$$R=12+4+14+15+10=55$$

利用递推公式 (9.8) 所作程序有如下输出:

```
THIS PROGRAM COMPUTES THE p-value FOR THE TWO-SAMPLE RANK
    SUM TEST
THIS PROGRAM WILL RUN FASTEST IF YOU DESIGNATE AS THE FIRST
```

233

```
    SAMPLE THE SAMPLE HAVING THE SMALLER SUM OF RANKS
ENTER THE SIZE OF THE FIRST SAMPLE
?  5
ENTER THE SIZE OF THE SECOND SAMPLE
?  10
ENTER THE SUM OF THE RANKS OF THE FIRST SAMPLE
?  55
The p-value IS 0.0752579
OK
```

□

应用 (9.8) 式计算 p 值的困难在于随样本容量增加，计算量会快速增长. 例如，当 $n = m = 20$，即使我们选择较小的秩和为统计量，由于所有秩的和为 $1 + 2 + \cdots + 40 = 820$，统计量的值可能为 410. 因此，为确定 p 值，有可能要计算 $20 \times 20 \times 410 = 164\ 000$ 个 $P_{n,m}(r)$ 个值. 这样，样本很大时，(9.8) 式可能不适用. 这种情况下，可以使用其他两种不同的近似方法 (a) 一个经典的方法就是近似 R 的分布，另一个是 (b) 模拟方法.

为使用经典的方法近似 p 值，我们利用事实：在原假设 H_0 下，$n + m$ 个值的排序是等可能的. 由此结论，容易证明

$$E_{H_0}[R] = n\frac{(n+m+1)}{2}$$

$$\mathrm{Var}_{H_0}(R) = nm\frac{(n+m+1)}{12}$$

可以证明，在原假设 H_0 下，当 n 和 m 很大时，R 近似正态分布. 因此，当 H_0 成立时，

$$\frac{R - n(n+m+1)/2}{\sqrt{nm(n+m+1)/12}} \text{ 近似为一个标准正态分布}$$

对一个正态随机变量 W，当 $r \leqslant E[W]$ 时，$P\{W \leqslant r\}$ 是 $P\{W \leqslant r\}$ 和 $P\{W \geqslant r\}$ 中较小的，否则 $P\{W \geqslant r\}$ 为较小值，因此，当 n 和 m 不太小时（两者都大于 7 就足够了），我们能近似检验结果 $R = r$ 的 p 值：

234

$$p\text{ 值} \approx \begin{cases} 2P\{Z < r^*\} & \text{如果} r \leqslant n\frac{(n+m+1)}{2} \\ 2P\{Z > r^*\} & \text{其他情况} \end{cases} \tag{9.9}$$

其中

$$r^* = \frac{r - \dfrac{n(n+m+1)}{2}}{\sqrt{\dfrac{nm(n+m+1)}{12}}}$$

且 Z 为一个标准正态随机变量.

例 9g 通过例 9g 的数据，我们来看经典的近似方法表现如何. 本例中，由于 $n=5$，$m=10$，我们有

$$\begin{aligned} p\text{ 值} &= 2P_{H_0}\{R \geqslant 55\} \\ &\approx 2P\left\{Z \geqslant \frac{55-40}{\sqrt{\frac{50\times 16}{12}}}\right\} \\ &= 2P\{Z \geqslant 1.8371\} = 0.066 \end{aligned}$$

可将其与确切值 0.075 相比较. □

两样本秩和检验的 p 值也可由模拟来近似. 要了解如何实现模拟过程，就要知道如果检验量 R 的观测值为 $R=r$，则有 p 值为

$$p\text{ 值} = 2\min\left(P_{H_0}\{R \geqslant r\}, P_{H_0}\{R \leqslant r\}\right)$$

现在，在原假设 H_0 下，设所有 $n+m$ 个数据值全不相同，则这些数据所有排列是等可能的，于是大小为 n 的第一组数据的秩与从 $1,2,\cdots,n+m$ 中随机选出 n 个值有相同的分布. 因此，在原假设 H_0 下，通过连续模拟产生整数 $1,2,\cdots,n+m$ 中大小为 n 的子集及确定子集中元素的和，可近似 R 的概率分布. $P_{H_0}\{R \leqslant r\}$ 的值可由模拟得到的小于或等于 r 的和的比例来近似，同时 $P_{H_0}\{R \geqslant r\}$ 的值可由模拟得到的大于或等于 r 的和的比例来近似.

上述分析假设所有 $n+m$ 个数据全不相同. 当它们中有某些值相同时，数据的秩为与其相等的所有数据的秩的平均. 例如，如果第一组数据为 2, 3, 4 且第二组数据为 3, 5, 7，则第一组数据的秩和为 $1+2.5+4=7.5$. 近似 p 值可使用 (9.9) 式的正态近似. 235

两样本问题的推广是多样本问题，其中有如下 m 组数据：

$$\begin{matrix} X_{1,1}, & X_{1,2}, & \cdots, & X_{1,n_1} \\ X_{2,1}, & X_{2,2}, & \cdots, & X_{2,n_2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m,1}, & X_{m,2}, & \cdots, & X_{m,n_m} \end{matrix}$$

要检验的原假设 H_0 为所有 $n=\sum_{i=1}^{m} n_i$ 个随机变量独立同分布. 两样本秩和检验的一个推广，称为多样本秩和检验（或通常的 Kruskal-Wallis 检验），可通过先将所有 n 个数据排序来获得. 然后，设 R_i，$i=1,\cdots,m$，为来自第 i 组数据的所有 n_i 个数据的秩和.（注意到符号 R_i 为秩和，而非前面所表示的单个数据的秩.）由于在

原假设 H_0 下，所有数据排序等可能出现（如果所有数据全不相同），则如前所述，我们有

$$E[R_i] = n_i\frac{(n+1)}{2}$$

利用上式，多样本秩和检验的检验量为

$$R = \frac{12}{n(n+1)}\sum_{i=1}^{m}\frac{[R_i - n_i(n+1)/2]^2}{n_i}$$

由于较小的 R 值表示对 H_0 的拟合较好，则统计量为 R 的检验，当 R 足够大时，拒绝原假设 H_0. 事实上，如果 R 的观测值为 $R = y$，则其 p 值为

$$p\text{ 值} = P_{H_0}\{R \geqslant y\}$$

此值可由 $n_1, \cdots, n_m$ 个数据得到其近似，且 R 近似服从自由度为 $m-1$ 的 χ^2 分布 [这就说明为什么在 R 的定义中包含了 $12/n(n+1)$ 这一项]. 因此，如果 $R = y$，

$$p\text{ 值} \approx P\{\chi_{m-1}^2 \geqslant y\}$$

可用模拟来计算 p 值.（见习题 14）

即使有数据相同，上述对 p 值的近似仍然适用. 计算 R 值时，个别数据的秩（如前所述）应为所有与其相等的数据的秩的平均.

236

9.4 验证非齐次泊松过程假设

考虑如下数学模型：假设每日到达系统的个数为一个非齐次泊松过程，且每天的到达过程独立，有相同但未知的强度函数.

为验证这一假设，我们观测系统 r 天，注意到达次数. 令 $N_i, i = 1, \cdots, r$，为第 i 天到达次数并注意到如果到达过程确为一个非齐次泊松过程，则这些量是独立且有相同均值的泊松随机变量. 尽管这一结果可以如例 9a 利用拟合优度方法进行检验，但我们给出的另出一种方法在有些情况下却更为有效. 此方法是基于泊松随机变量的均值和方差相同的事实. 因此，如果 N_i 确为来自泊松分布的样本，则样本均值

$$\overline{N} = \sum_{i=1}^{r}\frac{N_i}{r}$$

与样本方差

$$S^2 = \sum_{i=1}^{r}\frac{(N_i - \overline{N})^2}{r-1}$$

应大致相等. 综上分析, 我们建立假设检验如下:

H_0: N_i 为均值相同的独立泊松随机变量, 其检验统计量为

$$T = \frac{S^2}{\overline{N}} \tag{9.10}$$

由于 T 值很大或很小都将与原假设 H_0 不一致, 则当 $T = t$ 时, 其 p 值为

$$p \text{ 值} = 2\min\left(P_{H_0}\{T \leqslant t\}, P_{H_0}\{T \geqslant t\}\right)$$

然而, 由于 H_0 为未知均值的泊松分布, 我们目前不能计算上述概率; 相反, 我们必须先利用观测值估计均值. 利用估计 $\overline{N}$, 可知如果 $\overline{N}$ 的观测值为 $\overline{N} = m$, 则 p 值近似为

$$p \text{ 值} \approx 2\min\left(P_m\{T \leqslant t\}, P_m\{T \geqslant t\}\right)$$

237

其中 T 的定义如 (9.10) 式, $N_1, \cdots, N_r$ 是均值为 m 的独立泊松随机变量. 现通过模拟近似 $P_m\{T \leqslant t\}$ 和 $P_m\{T \geqslant t\}$. 即, 连续产生 r 个均值为 m 的独立泊松随机变量并计算 T 值. $T \leqslant t$ 的比例用来估计 $P\{T \leqslant t\}$, 同时 $T \geqslant t$ 的比例用来估计 $P\{T \geqslant t\}$.

如果上述 p 值很小, 则拒绝每日到达系统为一个非齐次泊松过程的原假设. 然而, 如果 p 值不是很小, 则只表明每天到达的次数服从一个泊松分布是一个可行的假设, 但其本身不能验证每天实际到达模型 (由非齐次泊松过程确定) 相同这一更强的假设. 为完成验证, 必须考虑观测的 r 天中每天到达的次数. 假设第 j 天 $(j = 1, \cdots, r)$ 到达次数为 $X_{j,1}, X_{j,2}, \cdots, X_{j,N_j}$. 如果到达过程确为一非齐次泊松过程, 则可以证明这 r 组到达次数中的每一个都可成为来自同一分布的一个样本. 即, 在原假设下, r 组数据 $X_{j,1}, \cdots, X_{j,N_j} (j = 1, \cdots, r)$ 为独立同分布的随机变量.

然而, 上述结论可以由 9.3 节中多样本秩和检验来进行检验. 即, 首先对所有 $N \equiv \sum_{j=1}^{r} N_j$ 个数据排序, 并令 R_j 为第 j 组所有 N_j 个数据的秩和. 可利用检验量

$$R = \frac{12}{N(N+1)} \sum_{j=1}^{r} \frac{\left(R_j - N_j \frac{(N+1)}{2}\right)^2}{N_j}$$

因为当 H_0 成立时, R 近似服从一个自由度为 $r-1$ 的 χ^2 分布. 因此, 如果观测值 $R = y$, p 值可近似为

$$\begin{aligned} p \text{ 值} &= 2\min\left(P_{H_0}\{R \leqslant y\}, P_{H_0}\{R \geqslant y\}\right) \\ &\approx 2\min\left(P\{\chi_{r-1}^2 \leqslant y\}, 1 - P\{\chi_{r-1}^2 \leqslant y\}\right) \end{aligned}$$

其中 χ^2_{r-1} 是自由度为 $r-1$ 的 χ^2 随机变量.（当然，我们也可以通过模拟来近似 p 值.）如果上述 p 值（包括前面考虑的 p 值）不是太小，则可知数据与每日到达系统为一个非齐次泊松过程的原假设一致.

238

技术注释　许多读者可能想知道为什么我们在计算 (9.11) 式中的 p 值时使用双边区域，而不使用在多样本秩和检验中的单边区域. 这是由于多样本秩和检验假设数据来自 m 个分布，且当这些分布相同时，R 很小，因此，单边概率计算 p 值是合适的. 然而，在检验一个周期性非齐次泊松过程时，我们既要检验第 i 天到达次数服从某个分布又要检验对所有天的分布相同. 即，我们不能同秩和检验一样，从开始就假设数据来自固定的几个分布. 因此，双边检验是适用的，原因是一个很小的 R 值可能表明一天的某个到达模型，也就是说，即使每天到达次数服从相同的泊松分布，每日到达时间仍可能不是独立同分布的. □

例 9h　考虑观测某电厂 5 天的日传送次数. 在这段时间内，每天传送次数如下：

$$18,\ 24,\ 16,\ 19,\ 25$$

同时假设这 102 次传送根据其到达时间排序，则每天传送的秩和为

$$1010,\ 960,\ 1180,\ 985,\ 1118$$

使用上述数据，检验每日传送到达过程为一个非齐次泊松过程的原假设.

首先检验每日传送次数的第一组数据为 5 个独立同分布的泊松随机变量. 样本均值和样本方差为

$$\overline{N}=20.4 \quad 和 \quad S^2=15.3$$

于是检验量的值为 $T=0.75$. 为检验 N_i 为独立泊松随机变量，及确定其近似 p 值，模拟产生 500 组 5 个均值为 20.4 的泊松随机变量并计算 $T=S^2/\overline{N}$. 模拟结果表明 p 值近似为 0.84，这清楚地表明每日传送次数为均值相同的独立泊松随机变量的假设与数据一致.

继续检验非齐次泊松过程原假设，计算检验统计量 R 值为 14.425. 由于自由度为 4 的 χ^2 随机变量大于等于 14.425 的概率为 0.006，于是 p 值为 0.012. 对这样小的p 值我们拒绝原假设. □

239

如果要检验假设每日到达过程为一个齐次泊松过程，应如前所述，首先检验原假设每天到达次数是独立同分布的泊松随机变量. 完成这一检验后，如果原假设成立，则同非齐次的情况一样，继续考虑 $N=\sum_{j=1}^{r} N_j$ 个到达次数的实际数据. 而且，我们知道，在齐次泊松过程中，已知一天之内的到达次数，到达时间是独立的在 $(0,T)$ 上的均匀分布，其中 T 为一天的长度. 而这一结果可以由 9.1 节的科尔莫戈

罗夫－斯米尔诺夫拟合优度检验进行检验. 即，如果到达为一个齐次泊松过程，N 个随机变量 $X_{j,i}$ 独立服从于 $(0,T)$ 上的均匀分布，$i=1,\cdots,N_j$，$j=1,\cdots,r$，其中 $X_{j,i}$ 表示在第 j 天的第 i 个到达时间. 因此，如果定义经验分布函数 F_e 为 N 个数据值中小于或等于 x 的比例，即，

$$F_e(x)=\sum_{j=1}^{r}\sum_{i=1}^{N_j}\frac{I_{j,i}}{N}$$

其中

$$I_{j,i}=\begin{cases}1 & X_{j,i}\leqslant x\\ 0 & \text{其他情况}\end{cases}$$

则检验统计量的值为

$$D=\max_{0\leqslant x\leqslant T}\left|F_e(x)-\frac{x}{T}\right|$$

只要确定检验统计量 D 的值，我们就可以如 9.1 节所示通过模拟方法得到 p 值.

如果非齐次泊松过程的原假设与数据一致，则我们面临的问题为估计过程的强度函数 $\lambda(t),0\leqslant t\leqslant T$.[在齐次情况下一个显然的估计为 $\lambda(t)=\hat{\lambda}/T$，其中 $\hat{\lambda}$ 是时间长度为 T 的一天之内到达次数的均值估计.] 为估计强度函数，排序 $N=\sum_{j=1}^{r}N_j$ 个日常到达时间. 设 $y_0=0$，且对 $k=1,\cdots,N$，令 y_k 为这 N 个到达次数中第 k 个最小的. 由于包含 r 天的时间间隔 (y_{k-1},y_k)，$k=1,\cdots,N$，有一次到达，因此关于 $\lambda(t)$ 的一个合理估计为：

$$\hat{\lambda}(t)=\frac{1}{r(y_k-y_{k-1})},\quad y_{k-1}<t<y_k$$

240

[要理解上述估计，注意到如果 $\hat{\lambda}(t)$ 为强度函数，则在时刻 t，$y_{k-1}<t\leqslant y_k$，每日到达次数的期望为

$$E[N(y_k)-N(y_{k-1})]=\int_{y_{k-1}}^{y_k}\hat{\lambda}(t)\mathrm{d}t=\frac{1}{r}$$

在 r 天内的时间间隔中到达的期望数为 1，与在该区间内实际观测的到达次数一致.]

习 题

1. 由遗传学中的孟德尔理论，某种豌豆花可能为白色、粉色或红色，其概率分别为 1/4, 1/2, 1/4. 要检验这一理论，研究 564 株豌豆，其中 141 株开白花，291 株开粉花，132 株开红花. 求此组数据的近似 p 值.

   (a) 利用 χ^2 近似；　(b) 利用模拟方法.

2. 为确定某一骰子是否均匀，记录掷 1000 次骰子的情况，出现点数 i，$i=1,2,3,4,5,6$ 的次数分别为 158，172，164，181，160，165. 用以下方法近似检验骰子是均匀的 p 值：

   (a) χ^2 近似；　　(b) 模拟方法.

3. 原假设是下列 10 个值为随机数，近似该假设的 p 值：0.12，0.18，0.06，0.33，0.72，0.83，0.36，0.27，0.77，0.74.

4. 原假设为下列 14 个点是来自 $(50,200)$ 上的均匀分布的样本，近似该假设的 p 值：

   164，142，110，153，103，52，174，88，178，184，58，62，132，128

5. 原假设为下列 13 个值是来自均值为 50 的指数分布，近似该假设的 p 值：

   241

   86，133，75，22，11，144，78，122，8，146，33，41，99

6. 检验下列数值来自参数为 $(8,p)$ 的二项分布，其中 p 未知. 近似求此检验的 p 值：

   6，7，3，4，7，3，7，2，6，3，7，8，2，1，3，5，8，7

7. 检验下列数值来自指数分布总体，近似其 p 值：122，133，106，128，135，126.

8. 为产生 n 个随机数的排序值，我们产生 n 个随机数并排序或分类. 另一方法是利用如下事实：已知泊松过程的第 $(n+1)$ 个事件出现在 t 时刻，前 n 个事件的发生时间的分布等同于在 $(0,t)$ 上均匀分布的 n 个随机变量排序后值的分布. 利用这一结果，可以解释为什么在下面算法中，$y_1,\cdots,y_n$ 定义为 n 个随机数排序后的值.

   生成 $n+1$ 个随机数 $U_1,\cdots,U_{n+1}$

   $$X_i=-\ln U_i \quad i=1,\cdots,n+1$$

   $$t=\sum_{i=1}^{n+1}X_i, \quad c=\frac{1}{t}$$

   $$y_i=y_{i-1}+cX_i, \quad i=1,\cdots,n(\text{其中 } y_0=0)$$

9. 产生 10 个均值为 1 的独立的指数随机变量. 则由科尔莫戈罗夫－斯米尔诺夫检验量，对检验数据的确是来自一个均值为 1 的指数分布，近似求此检验的 p 值.

10. 设计一试验比较使电线抗腐蚀的两个处理方法，并对几段电线采用其中一种方法进行处理，得到如下数据（表示以千分之一英寸为单位腐蚀点的最大深度）：

    方法 1：65.2　67.1　69.4　78.4　74.0　80.3

    方法 2：59.4　72.1　68.0　66.2　58.5

    检验两处理有相同结果的假设，计算这组数据的准确 p 值.

11. 通过如下两种方法计算习题 10 中的近似 p 值：

    (a) 正态近似；　　(b) 模拟方法.

12. 选择 14 个大小相当的城市研究交通安全问题. 随机选取 7 个城市在报纸上开展为期 1 个月的交通安全宣传，活动结束后下 1 个月报道的交通事故如下： 242

    处理组：19　31　39　45　47　66　75

    控制组：28　36　44　49　52　72　72

    检验报道没有效果的原假设，确定准确的 p 值.

13. 用下列方法近似习题 12 中的 p 值：

    (a) 正态近似；　　(b) 模拟方法.

14. 说明在多样本问题中如何应用模拟近似 p 值，即，检验 m 个样本来自相同概率分布的情况.

15. 考虑下面来自三个样本的数据：

    样本 1：121　144　158　169　194　211　242

    样本 2：99　128　165　193　242　265　302

    样本 3：129　134　137　143　152　159　170

    检验所有数据来自同一概率分布，计算近似 p 值：

    (a) 利用 χ^2 近似；　　(b) 利用模拟方法.

16. 在一个 8 天间隔内每日到达次数如下：

    122, 118, 120, 116, 125, 119, 124, 130

    你认为每日到达次数是否独立同分布于非齐次泊松过程?

17. 一长度为 100 的间隔，在如下时间有 18 次到达：

    12, 20, 33, 44, 55, 56, 61, 63, 66, 70, 73, 75, 78, 80, 82, 85, 87, 90

    检验到达过程为一个（齐次）泊松过程，近似其 p 值.

243

参考文献

[1] Diaconis, P., and B. Efron, *Computer Intensive Methods in Statistics,Sci. Am.*, **248**(5), 96-109, 1983.

[2] Fishman, G. S., *Concepts and Methods in Discrete Event Digital Simulations.* Wiley, New York, 1973.

[3] Kendall, M., and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. MacMillan, New York, 1979.

[4] Mihram, G. A., *Simulation—Statistical Foundations and Methodology.* Academic Press, New York, 1972.

[5] Sargent, R. G., *A Tutorial on Validation and Verification of Simulation Models, Proc. 1988 Winter Simulation Conf.*, San Diego, pp.33-39, 1988.

[6] Schruben, L. W.,*Establishing the Credibility of Simulations, Simulation*, **34**, 101-105, 1980.

第10章　MCMC 方法

引　　言

一般来说，模拟一个其分量为相关随机变量的随机向量 $\boldsymbol{X}$ 是件困难的事. 在这一章里，我们介绍一强有力的方法来生成分布近似为 $\boldsymbol{X}$ 的分布的随机向量. 这种方法被称为MCMC 方法，它仅需要一个由乘积常数确定的 $\boldsymbol{X}$ 的密度函数 (或分布列)，而我们将会看到, 这一点在应用中非常重要.

在 10.1 节中，我们介绍一些后面需要用到的关于马氏链的结论. 在 10.2 节中给出 Hastings-Metropolis 算法来构造一个将某一特定概率密度函数作为其极限分布的马氏链. 这种算法的一个特例，被称为吉布斯抽样会在 10.3 节中讨论. 这种抽样法可能是最常用的MCMC 方法. 在 10.4 节中将讨论，把前面的方法应用到确定性优化问题，也就是熟知的模拟退火方法. 10.5 节中，我们给出抽取重要再抽样技术 (SIR)，它尽管不是一个严格的MCMC 方法，但也能够近似地生成密度函数为常数乘积的随机向量.

10.1　马　氏　链

考虑一组随机变量 $X_0, X_1, \cdots$. 这里把 X_n 解释为“在时间点 n”系统的状态，并假设 X_n 可能取值的集合，即系统可能达到的状态为集合 $1, \cdots, N$. 如果存在一组数 $P_{ij}, i, j = 1, \cdots, N$，使得过程无论什么时候处于状态 i，也不管前面的状态如何，其下一个状态是 j 的概率为 P_{ij}，我们就称 $\{X_n, n \geqslant 0\}$ 构成一个转移概率为 $P_{ij}, i, j = 1, \cdots, N$ 的马氏链. 因为过程离开状态 i 之后必定会处于某一个状态，所以这些转移概率满足

245

$$\sum_{j=1}^{N} P_{ij} = 1, \quad i = 1, \cdots, N$$

如果对于每一对状态 i 和 j，从状态 i 出发，过程会转移到状态 j 的概率大于零，则这样的马氏链称为是不可约的. 对于一个不可约马氏链，令 π_j 表示经过相当长的一段运行时间过程处于状态 j 的时间的比例 (可以证明 π_j 为一常数，以概率 1 存在，并与初始状态独立). 可以证明 $\pi_j, j = 1, \cdots, N$ 为下面线性等式组的唯一

解：

$$\pi_j = \sum_{i=1}^{N} \pi_i P_{ij}, \quad j = 1, \cdots, N$$
$$\sum_{j=1}^{N} \pi_j = 1 \tag{10.1}$$

注　关于等式组 (10.1)，我们有一个启发性的解释. 因为 π_i 是马氏链处于状态 i 的比例，而从 i 转移到 j 的概率为 P_{ij}，那么 $\pi_i P_{ij}$ 就是马氏链从状态 i 转移到 j 的时间的比例. 因此，(10.1) 中的上面的等式就表示，马氏链处于状态 j 的时间所占的比例等于马氏链从状态 i 转移到 j 所占的比例对 i 求和. (10.1) 中的下面的等式当然就是表示，马氏链处于状态 j 的时间所占的比例对 j 求和，必定等于 1. □

上述 $\{\pi_j\}$ 被称为马氏链的*平稳概率*. 如果马氏链的初始状态分布为 $\{\pi_j\}$ 的话，则 $P\{X_n = j\} = \pi_j$，对所有的 n, j 均成立（见习题 1）.

马氏链的一个重要性质是对于状态空间中的任意函数 h，有

246

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} h(X_i) = \sum_{j=1}^{N} \pi_j h(j) \tag{10.2}$$

以概率 1 成立. 这是由于如果 $p_j(n)$ 是在时刻 $1, \cdots, n$ 之间，马氏链处于状态 j 的时间比例，则

$$\frac{1}{n} \sum_{i=1}^{n} h(X_i) = \sum_{j=1}^{N} h(j) p_j(n) \to \sum_{j=1}^{N} h(j) \pi_j$$

π_j 可以解释为马氏链处于 j 的极限概率. 为给出此解释的精确条件，我们首先需给出非周期的马氏链的定义.

定义　*一个不可约的马氏链被称为非周期的，如果对某个 $n \geqslant 0$ 和某个状态 j，有*

$$P\{X_n = j | X_0 = j\} > 0 \text{ 和 } P\{X_{n+1} = j | X_0 = j\} > 0$$

可以证明，如果一马氏链是不可约且非周期的，则

$$\pi_j = \lim_{n\to\infty} P\{X_n = j\}, \; j = 1, \cdots, N$$

在有些时候，有比求等式组 (10.1) 的解更好的方法来获得平稳概率. 假设我们可以找到一些正数 $x_j, j = 1, \cdots, N$ 使得下式成立

$$x_i P_{ij} = x_j P_{ji}, \text{ 对 } i \neq j, \; \sum_{j=1}^{N} x_j = 1$$

则将上述前一个等式对所有的状态 i 求和得到

$$\sum_{i=1}^{N} x_i P_{ij} = x_j \sum_{i=1}^{N} P_{ji} = x_j$$

又因为 $\{\pi_j, j = 1, \cdots, N\}$ 是 (10.1) 的唯一解，则意味着

$$\pi_j = x_j$$

当对所有的 $i \neq j$, $\pi_i P_{ij} = \pi_j P_{ji}$ 都成立时，马氏链被称作时间可逆的，因为在此条件下，我们可以证明，如果用概率 $\{\pi_j\}$ 来做初始状态的话，那么从任意时刻开始，以相反的时间方向来看，状态序列仍旧是一转移概率为 P_{ij} 的马氏链.

现在想要生成具有概率分布函数 $P\{X = j\} = p_j, j = 1, \cdots, N$ 的随机变量 X 的值. 假如能够生成一个具有极限概率 $p_j, j = 1, \cdots, N$ 的不可约非周期的马氏链，那么我们就能够通过运行该马氏链，n 步 (n 很大)，来获得 X_n 的值，并将这个值近似作为想要生成的随机变量值. 另外，如果我们的目标是为了能够估计 $E[h(X)] = \sum_{j=1}^{N} h(j)p_j$ 而生成许多分布为 $p_j, j = 1, \cdots, N$ 的随机变量，那么我们同样可以使用估计量 $\frac{1}{n}\sum_{i=1}^{n} h(X_i)$ 来估计它. 然而，由于初始状态的选取对前期的马氏链状态有很强的影响，一般在应用中常用的方法是适当地选取某个 k，然后舍去前 k 个状态. 也就是说，使用估计量 $\frac{1}{n-k}\sum_{i=k+1}^{n} h(X_i)$. 到底使用多大的 k 值很难精确地知道 [但是水平高的读者应该可以从 Aarts and Korst (1989) 上找到一些相关的有用的结果]，并且通常可以用直观上给出的值（通常来说这样做能有比较好的效果，因为无论取什么样的值都能够有很好的收敛）.

247

一个重要的问题是如何用马氏链来估计一个估计量的均方误差. 也就是说，如果令 $\hat{\theta} = \frac{1}{n-k}\sum_{i=k+1}^{n} h(X_i)$，我们如何来估计

$$\text{MSE} = E\left[\left(\hat{\theta} - \sum_{j=1}^{N} h(j)p_j\right)^2\right]$$

一种方法是使用如下的组均值方法. 将 $n-k$ 个生成的状态分成 s 组，每组容量为 r，其中 $s = (n-k)/r$ 为一整数，并令 $Y_j, j = 1, \cdots, s$ 为第 j 组的平均. 也就是说，

$$Y_j = \frac{1}{r} \sum_{i=k+(j-1)r+1}^{k+jr} h(X_i),\ j = 1, \cdots, s$$

现把 $Y_j, j = 1, \cdots, s$ 看作是方差为 σ^2 的独立同分布的随机变量，然后使用它们的样本方差 $\hat{\sigma}^2 = \sum_{j=1}^{s}(Y_j - \overline{Y})^2/(s-1)$ 作为 σ^2 的估计. 则 MSE 的估计为 $\hat{\sigma}^2/s$.

r 值的选取依赖于模拟生成的马氏链. $X_i, i \geqslant 1$ 之间越接近，将会越接近独立同分布，那么要选择更小的 r 值. 在下面的两节当中，我们会介绍在给定一组正数 $b_j, j = 1, \cdots, N$ 时，我们如何来构造极限概率为 $\pi_j = b_j / \sum_{i=1}^{N} b_i, j = 1, \cdots, N$ 的马氏链.

10.2 Hastings-Metropolis 算法

令 $b(j), j = 1, \cdots, m$ 为正数，且 $B = \sum_{j=1}^{m} b(j)$. 假设 m 很大且 B 很难计算，而我们想要模拟一随机变量（或一列随机变量），它具有如下概率分布函数

248

$$\pi(j) = b(j)/B, \ j = 1, \cdots, m$$

一种生成极限分布为 $\pi(j), j = 1, \cdots, m$ 的随机变量的方法是找一个容易计算的马氏链且其极限概率为 π_j. Hastings-Metropolis *算法* 提供了一种方法来完成这个任务. 它构造出一个具有所要求的极限概率且时间可逆的马氏链，具体方法如下.

令 $\boldsymbol{Q}$ 是状态为整数 $1, \cdots, m$ 的不可约马氏链的转移概率矩阵，其第 i 行第 j 列的元素为 $q(i,j)$. 现定义一马氏链 $\{X_n, n \geqslant 0\}$ 如下. 当 $X_n = i$ 时，生成一随机变量 X 使得 $P\{X = j\} = q(i,j)$, $j = 1, \cdots, m$. 如果 $X = j$，那么 X_{n+1} 就以 $\alpha(i,j)$ 的概率置为 j，以 $1 - \alpha(i,j)$ 的概率置为 i. 在这些条件下，容易证明状态序列将构成一具有如下转移概率 $P_{i,j}$ 的马氏链：

$$P_{i,j} = q(i,j)\alpha(i,j), \quad \text{如果} j \neq i$$
$$P_{i,i} = q(i,i) + \sum_{k \neq i} q(i,k)(1 - \alpha(i,k))$$

这个马氏链是时间可逆的且有平稳概率 $\pi(j)$，如果

$$\pi(i)P_{i,j} = \pi(j)P_{j,i} \ \text{对} j \neq i$$

或等价地

$$\pi(i)q(i,j)\alpha(i,j) = \pi(j)q(j,i)\alpha(j,i)$$

如果我们取

$$\alpha(i,j) = \min\left(\frac{\pi(j)q(j,i)}{\pi(i)q(i,j)}, 1\right) = \min\left(\frac{b(j)q(j,i)}{b(i)q(i,j)}, 1\right) \tag{10.3}$$

则容易验证上式成立. [为验证这一点，注意到，如果 $\alpha(i,j) = \pi(j)q(j,i)/\pi(i)q(i,j)$，那么 $\alpha(j,i) = 1$，反之亦然.]

读者应当注意的是，由于 $b(j)$ 的值已足够，故 B 的值在定义马氏链时是不需要的. 而且，一种很常见的情况是 $\pi(j), j=1,\cdots,m$ 不仅是平稳概率，而且是极限概率.（事实上，一个充分条件是对于某个 i 有 $p_{i,i}>0$.）

下面总结上述用于生成极限概率为 $\pi(j)=b(j)/B,\ j=1,\cdots,m$ 的时间可逆马氏链的 Hastings-Metropolis 算法.

1. 选择一个转移概率为 $q(i,j), i,j=1,\cdots,m$ 的不可约马氏转移概率矩阵 $\boldsymbol{Q}$. 并选择某个 1 到 m 之间的整数 k.

249

2. 令 $n=0$ 及 $X_0=k$.
3. 生成一随机变量 X 使得 $P\{X=j\}=q(X_n,j)$，并生成一随机数 U.
4. 若 $U<[b(X)q(X,X_n)]/[b(X_n)q(X_n,X)]$，则 $NS=X$；否则 $NS=X_n$.
5. $n=n+1, X_n=NS$.
6. 转到第 3 步.

例 10a 假设我们想要生成来自一个大且复杂的组合集合 l 中的一个元素. 举例来说，对于一给定的常数 a, l 可能是 $(1,\cdots,n)$ 的所有满足 $\sum_{j=1}^{n} jx_j>a$ 的排列 $(x_1,\cdots,x_n)$ 的集合；或者 l 可能是一给定图的满足如下性质的所有子图的集合：对任意的一组顶点 i 和 j，在子图上存在唯一的从 i 到达 j 的路径（这样的子图被称为树.）

为了实现我们的目标，我们使用 Hastings-Metropolis 算法. 首先假设可以定义 l 的相邻元素的概念，则我们通过在 l 的每对相邻元素之间设置一弧来构造一个顶点集合为 l 的图. 举例来说，如果 l 是满足 $\sum_{j=1}^{n} jx_j>a$ 的排列 $(x_1,\cdots,x_n)$ 的集合，那么若一个排列是由另一个排列的位置中的两个相交换后得到的，那么我们就可以定义这样的两个排列为相邻的. 也就是说 $(1,2,3,4)$ 和 $(1,2,4,3)$ 是相邻的，而 $(1,2,3,4)$ 和 $(1,3,4,2)$ 不相邻. 如果 l 是一树的集合，那么如果一个树的弧中只有一个不是另外一个树的弧，那么我们可以称这样的两个树为相邻的.

在假设了这样的相邻元素概念之后，我们定义 q 个转移概率等式如下. 用 $N(s)$ 来定义 s 的相邻元素的集合，并且 $|N(s)|$ 等于集合 $N(s)$ 中元素的个数，令

$$q(s,t)=\frac{1}{|N(s)|},\ \text{如果 } t\in N(s)$$

也就是说，从 s 出发的下一个状态等可能地为其所有的相邻元素. 因为想要的马氏链的极限概率为 $\pi(s)=C$，那么 $\pi(s)=\pi(t)$，并且有

$$\alpha(s,t)=\min(|N(s)|/|N(t)|,1)$$

即，如果马氏链现在的状态为 s，其相邻元素中的一个将被随机地选取，记为 t. 如果 t 比 s 有更少的相邻元素（用图论的语言，就是顶点 t 的度比 s 的小），则下一

个状态就是 t. 若不是这样的话, 生成一个随机数 U, 如果 $U < |N(s)|/|N(t)|$, 则下

250

一个状态就是 t, 否则的话就是 s. 马氏链的极限概率就是 $\pi(s) = 1/|l|$. □

10.3　吉布斯抽样

应用 Hastings-Metropolis 算法最广泛的是吉布斯抽样. 设随机向量 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的概率分布函数（或连续型随机向量的概率密度函数）为 $p(\boldsymbol{x})$, 其中 $p(\boldsymbol{x})$ 仅需确定一个乘积常数, 并假定我们要生成一个与 $\boldsymbol{X}$ 同分布的向量. 即, 要产生一个密度函数为

$$p(\boldsymbol{x}) = Cg(\boldsymbol{x})$$

的随机向量, 其中 $g(\boldsymbol{x})$ 已知, 而 C 未知. 利用吉布斯抽样假设对任意 i 和 x_j, $j \neq i$, 产生一分布函数为

$$P\{X = x\} = P\{X_i = x | X_j = x_j, j \neq i\} \tag{10.4}$$

的随机变量 X.

对状态为 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 的马氏链应用 Hastings-Metropolis 算法, 其转移概率定义如下. 如果当前状态为 $\boldsymbol{x}$, 从 $1, \cdots, n$ 中等可能选取一个坐标. 如果选取坐标 i, 则可得到一随机变量 X, 其概率分布函数由 (10.4) 式给出, 且如果 $X = x$, 则 $\boldsymbol{y} = (x_1, \cdots, x_{i-1}, x, x_{i+1}, \cdots, x_n)$ 为下一个状态的坐标. 换句话说, 已知 $\boldsymbol{x}$ 和 $\boldsymbol{y}$, 吉布斯抽样利用 Hastings-Metropolis 算法, 有

$$q(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n} P\{X_i = x | X_j = x_j, j \neq i\} = \frac{p(\boldsymbol{y})}{nP\{X_j = x_j, j \neq i\}}$$

由于极限分布函数为 p, 则根据 (10.3) 式, 向量 $\boldsymbol{y}$ 作为新状态的概率为

$$\begin{aligned}
\alpha(\boldsymbol{x}, \boldsymbol{y}) &= \min\left(\frac{p(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x})}{p(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y})}, 1\right) \\
&= \min\left(\frac{p(\boldsymbol{y})p(\boldsymbol{x})}{p(\boldsymbol{x})p(\boldsymbol{y})}, 1\right) \\
&= 1
\end{aligned}$$

因此, 利用吉布斯抽样得到的坐标总可以作为链的下一步状态.

例 10b　设要产生以原点为圆心, 以 1 为半径的圆上的 n 个随机点, 约束条件为任意两个点的距离均大于 d, 且

251

$$\beta = P\{任意两个点的距离均大于d\}$$

为一个很小的正数. (如果 β 不是很小, 则继续产生圆上 n 个随机点, 直到首次出现任意两个点的距离均大于 d 时停止.) 这一过程可由吉布斯抽样完成, 其中起始点为圆上 n 个点 $x_1,\cdots,x_n$, 使得任意两个点的距离均大于 d. 然后产生随机数 U 并设 $I=\text{Int}(nU)+1$. 同时产生圆上的一个随机点. 如果这个点与除 x_I 外的其他 $n-1$ 个点的距离均大于 d, 则用这个点代替 x_I; 否则, 产生一个新的点并重复上述过程. 多次重复这一过程后, 得到的 n 个点将近似服从所要求的分布. □

例 10c 排队网络 设 r 个个体在 $m+1$ 个排队站点之间移动, 并设 $X_i(t)$ 为 t 时刻站点 i 上的个体数. 如果

$$p(n_1,\cdots,n_m)=\lim_{t\to\infty}P\{X_i(t)=n_i, i=1,\cdots,m\}$$

且假设服务次数服从指数分布, 则有

$$p(n_1,\ldots,n_m)=C\prod_{i=1}^{m}P_i(n_i),\quad \text{如果}\sum_{i=1}^{m}n_i\leqslant r$$

其中 $P_i(n)$, $n\geqslant 0$ 为概率分布函数, $i=1,\cdots,m$. 这样的联合概率分布函数称为有*乘积形式*.

虽然给出有上述乘积形式的 $p(n_1,\cdots,n_m)$ 及获得分布列 P_i 往往相对比较直接, 但是要明确计算常数 C 仍是困难的. 即使

$$C\sum_{\boldsymbol{n}:s(\boldsymbol{n})\leqslant r}\prod_{i=1}^{m}P_i(n_i)=1$$

其中 $\boldsymbol{n}=(n_1,\cdots,n_m)$, $s(\boldsymbol{n})=\sum_{i=1}^{m}n_i$, 这一结论也难以利用. 这是由于对所有满足 $\sum_{i=1}^{m}n_i\leqslant r$ 的非负整数向量 $\boldsymbol{n}$ 求和 [有 $\binom{r+m}{m}$ 个这样的向量], 当 m 和 r 不是很大时, 已经相当大了.

为解决 $p(n_1,\cdots,n_m)$, 即计算 C 的困难, 另一方法是应用吉布斯抽样来产生一列近似服从 p 的序列.

252

首先, 如果 $N=(N_1,\cdots,N_m)$ 有联合分布函数 p, 则对于 $n=0,\cdots,r-\sum_{k\neq i}n_k$,

$$\begin{aligned}
&P\{N_i=n|N_1=n_1,\cdots,N_{i-1}=n_{i-1},N_{i+1}=n_{i+1},\cdots,N_m=n_m\}\\
&=\frac{p(n_1,\cdots,n_{i-1},n,n_{i+1},\cdots,n_m)}{\sum\limits_j p(n_1,\cdots,n_{i-1},j,n_{i+1},\cdots,n_m)}\\
&=\frac{P_i(n)}{\sum\limits_j P_i(j)}
\end{aligned}$$

其中上述和式为对所有 $j=0,\cdots,r-\sum_{k\neq i}n_k$ 的累加. 换句话说, 给定 $N_j(j\neq i)$ 下 N_i 的条件分布与一概率分布为 P_i 的随机变量在给定其值小于或等于 $r-\sum_{j\neq i}N_j$ 下的条件分布相同.

于是, 我们可以产生极限概率分布列为 $p(n_1,\cdots,n_m)$ 的马氏链的值如下:

1. 设 $(n_1,\cdots,n_m)$ 为任意非负整数满足 $\sum_i n_i\leqslant r$.
2. 产生 U 并设 $I=\text{Int}(mU+1)$.
3. 如果 $I=i$, 则设 X_i 有分布列 P_i 并产生一个随机变量 N, 其分布为以 $X_i\leqslant r-\sum_{j\neq i}n_j$ 为条件的 X_i 的条件分布.
4. 设 $n_i=N$ 且回到 2.

$(n_1,\cdots,n_m)$ 的一列连续值就构成一极限分布为 p 的马氏链的一系列状态. 所有与 p 有关的我们感兴趣的量都可由这组值估计. 例如, 这些向量第 j 个坐标的平均值收敛到在站点 j 的个体数的均值, 第 j 个坐标小于 k 的向量的比例收敛到在站点 j 上的个体数小于 k 的极限概率, 诸如此类. □

例 10d　设 X_i, $i=1,\cdots,n$ 是独立的随机变量, 且 X_i 服从参数为 λ_i, $i=1,\cdots,n$ 的指数分布. 设 $S=\sum_{i=1}^n X_i$, c 为一个大的正常数, 现想产生满足条件 $S>c$ 的随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$, 即, 要产生一个分布函数为

$$f(x_1,\cdots,x_n)=\frac{1}{P\{S>c\}}\prod_{i=1}^n\lambda_i\mathrm{e}^{-\lambda_i x_i},\quad \text{如果}\sum_{i=1}^n x_i>c$$

的随机向量. 要使上述要求容易实现, 首先以初始向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$ 开始, 且向量满足 $x_i>0$, $i=1,\cdots,n$, 及 $\sum_{i=1}^n x_i>c$, 然后, 产生一个随机数 U 并令 $I=\text{Int}(nU+1)$. 假设 $I=i$. 现要产生一个在条件 $X+\sum_{j\neq i}x_j>c$ 下服从参数为 λ_i 的指数随机变量 X, 即要生成一个在其值大于 $c-\sum_{j\neq i}x_j$ 条件下的 X. 由于以其大于一个正的常数为条件的指数随机变量服从的分布为这个常数加上此指数, 于是, 我们应生成一个参数为 λ_i 的指数随机变量 Y (设 $Y=-1/\lambda_i\ln U$), 并令

253

$$X=Y+\left(c-\sum_{j\neq i}x_j\right)^+$$

其中当 $b>0$ 时, b^+ 等于 b, 否则等于 0. 然后 x_i 的值重设为 X, 开始再次重复此算法. □

现假设要估计

$$\alpha=P\{h(\boldsymbol{X})>a\}$$

其中 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是一个随机向量, h 是 $\boldsymbol{X}$ 的一个任意函数, 且 α 很小. 由于 $h(\boldsymbol{X})$ 的一个生成值几乎总是小于 α, 如果直接使用吉布斯抽样方法产生分布列

收敛到 $\boldsymbol{X}$ 的分布，则要获得误差相对 α 很小的估计，将花费大量的时间. 因此，考虑下列方法.

首先，注意到对于 $-\infty = a_0 < a_1 < a_2 < \cdots < a_k = a$,

$$\alpha = \prod_{i=1}^{k} P\{h(\boldsymbol{X}) > a_i | h(\boldsymbol{X}) > a_{i-1}\}$$

于是，可以通过求得 $P\{h(\boldsymbol{X}) > a_i|h(\boldsymbol{X}) > a_{i-1}\}, i = 1, \cdots, k$, 来获得 α 的估计. 为使这种方法更有效，取 a_i, $i = 1, \cdots, k$ 使得 $P\{h(\boldsymbol{X}) > a_i|h(\boldsymbol{X}) > a_{i-1}\}$ 为大小适中的值.

为估计 $P\{h(\boldsymbol{X}) > a_i|h(\boldsymbol{X}) > a_{i-1}\}$，使用吉布斯抽样，其步骤如下：

1. 设 $J = N = 0$.
2. 选择使得 $h(\boldsymbol{x}) > a_{i-1}$ 的向量 $\boldsymbol{x}$.
3. 产生随机数 U 并设 $I = \text{Int}(nU) + 1$.
4. 如果 $I = k$, 则由给定 $X_j = x_j$, $j \neq k$ 下 X_k 的条件分布产生 X.
5. 如果 $h(x_1, \cdots, x_{k-1}, X, x_{k+1}, \cdots, x_n) \leqslant a_{i-1}$, 回到 4.
6. $N = N + 1$, $x_k = X$.
7. 如果 $h(x_1, \cdots, x_a) > a_i$, 则 $J = J + 1$.
8. 回到 3.

254

J 与 N 最后值的比即为 $P\{h(\boldsymbol{X}) > a_i|h(\boldsymbol{X}) > a_{i-1}\}$ 的估计.

例 10e 设在例 10d 的排队网络模型中，服务员 i 的服务时间服从参数为 μ_i, $i = 1, \cdots, m+1$ 的指数分布，且当一位顾客由服务员 i 完成服务后（与所有其他服务员独立），此顾客以概率 P_{ij} 到服务员 j 处排队（或当此服务员空闲时即时得到服务），其中 $\sum_{j=1}^{m+1} P_{ij} = 1$. 可以证明，得到服务员 $1, \cdots, m$ 服务的顾客数的极限概率分布函数为

$$p(n_1, \cdots, n_m) = C \prod_{j=1}^{m} \left(\frac{\pi_j \mu_{m+1}}{\pi_{m+1} \mu_j} \right)^{n_j}$$

其中 π_j, $j = 1, \cdots, m+1$ 是转移概率为 P_{ij} 的马氏链的平稳概率. 即，它们是

$$\pi_j = \sum_{i=1}^{m+1} \pi_i P_{ij}$$

$$\sum_{j=1}^{m+1} \pi_j = 1$$

的唯一解. 如果对服务员重新编号使得 $\max(\pi_j/\mu_j) = \pi_{m+1}/\mu_{m+1}$，则设 $a_j = \pi_j\mu_{m+1}/\pi_{m+1}\mu_j$，对于 $\sum_{j=1}^{m} n_j \leqslant r$，

$$p(n_1, \cdots, n_m) = C\prod_{j=1}^{m}(a_j)^{n_j}$$

其中 $0 \leqslant a_j \leqslant 1$. 由此易知，已知在其他 $m-1$ 个服务站上的顾客数为 $n_j(j \neq i)$ 条件下，在服务员 i 下的顾客数的条件分布与 -1 加上一个参数为 $1-a_i$ 的几何随机变量的条件分布相同，后者的条件分布为给定这一几何变量小于或等于 $r+1-\sum_{j\neq i} n_j$ 条件下求得的.

如果对所有 j, π_j 和 μ_j 均为常数，则在给定除服务员 $m+1$ 外的其他服务员上的顾客数为 $n_j(j \neq i)$ 条件下，在服务员 i 下的顾客数的条件分布为在 $0, 1, \cdots, r-\sum_{j\neq i} n_j$ 上的离散均匀分布. 设在这种情况下，$m = 20$，$r = 100$，要估计在服务员 1 下的顾客数 (记为 X_1) 大于 18 的极限概率. 设 $t_0 = -1$，$t_1 = 5$，$t_2 = 9$，$t_3 = 12$，$t_4 = 15$，$t_5 = 17$，$t_6 = 18$，利用吉布斯抽样连续估计 $P\{X_1 > t_i | X_1 > t_{i-1}\}$，$i = 1, 2, 3, 4, 5, 6$. 通过从向量 $n_1, \cdots, n_{20}$ 开始，其中 $n_1 > 15$ 且 $s = \sum_{i=1}^{20} n_i \leqslant 100$，估计 $P\{X_1 > 17 | X_1 > 15\}$. 接着，产生一个随机数 U 并设 $I = \text{Int}(20U+1)$. 然后，产生第二个随机数 V. 如果 $I = 1$，则重设 n_1 为

255

$$n_1 = \text{Int}((85 - s + n_1)V) + 16$$

如果 $I \neq 1$，则重设 n_1 为

$$n_1 = \text{Int}((101 - s + n_1)V)$$

之后，重复上述算法，并用 $n_1 > 17$ 的比例估计 $P\{X_1 > 17 | X_1 > 15\}$. □

将一个小概率写成几个大小适中的条件概率的乘积，之后依次估计每个条件概率的想法并不需要用到吉布斯抽样. 而 Hastings-Metropolis 算法的另一种变化可能更为合适. 我们由前面使用重要抽样处理的例 8v 来说明.

例 10f　设要估计满足条件 $t(\boldsymbol{x}) > a$ 的关于 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 的排列个数，其中 $t(x) = \sum_{j=1}^{n} jx_j$ 且 a 使得排列数相对 $n!$ 较小. 如果设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为一排列，其取法在 $n!$ 种排列中是等可能的，并设

$$\alpha = P\{T(\boldsymbol{X}) > a\}$$

则 α 很小，且感兴趣的值为 $\alpha n!$. 设 $0 = a_0 < a_1 < \cdots < a_k = a$，则我们有

$$\alpha = \prod_{i=1}^{k} P\{T(\boldsymbol{X}) > a_i \mid T(\boldsymbol{X}) > a_{i-1}\}$$

要估计 $P\{T(\boldsymbol{X}) > a_i|T(\boldsymbol{X}) > a_{i-1}\}$，同例 10a 和 例 10b 一样，我们使用 Hastings-Metropolis 算法产生一马氏链，其极限分布为

$$\pi(\boldsymbol{x}) = \frac{1}{N_{i-1}}, \quad 如果\ T(\boldsymbol{x}) > a_{i-1}$$

其中 N_{i-1} 为使得 $T(\boldsymbol{x}) > a_{i-1}$ 的 $\boldsymbol{x}$ 的排列数. 所生成的满足 $T(\boldsymbol{x}) > a_i$ 的马氏链处于状态 $\boldsymbol{x}$ 的比例即为 $P\{T(\boldsymbol{X}) > a_i \mid T(\boldsymbol{X}) > a_{i-1}\}$ 的估计. □ 256

在许多应用中，都比较容易确认吉布斯抽样所需条件分布的形式.

例 10g 设对某非负函数 $h(y,z)$，非负随机变量 X，Y，Z 的联合密度为

$$f(x,y,z) = Cx^{y-1}(1-x)^{zy}h(y,z), \quad 0 < x < 0.5$$

则已知 $Y=y$ 和 $Z=z$ 下 X 的条件密度为

$$f(x|y,z) = \frac{f(x,y,z)}{f_{Y,Z}(y,z)}$$

由于 y 和 z 是固定的且 x 为此条件密度的自变量，于是上式可写为

$$f(x|y,z) = C_1 f(x,y,z)$$

其中 C_1 不依赖于 x. 因此，有

$$f(x|y,z) = C_2 x^{y-1}(1-x)^{zy}, \quad 0 < x < 0.5$$

其中 C_2 不依赖于 x. 可以发现这一条件密度与参数为 y 和 $zy+1$ 的 β 随机变量在区间 $(0,0.5)$ 上的条件密度相同.

相对于根据当前情况选择一个随机坐标，吉布斯抽样也可以考虑一列坐标. 即，在第一次重复中，可以设 $I=1$，在下次重复中，设 $I=2$，之后 $I=3$，依此类推，直到第 n 次重复，其中 $I=n$. 之后再重新开始. 下面的例子将说明这一过程，此例将模拟棒球比赛中，两名最好击球员击出本垒打的次数.

例 10h 设 $N_1(t)$ 为一个棒球赛季前 $100t\%$ 的赛程中球员巴里 · 邦兹本垒打的次数，$0 \leqslant t \leqslant 1$，类似地，设 $N_2(t)$ 为肯 · 格里菲的本垒打数.

设存在随机变量 W_1 和 W_2，满足在给定 $W_1 = w_1$，$W_2 = w_2$ 下，$\{N_1(t), 0 \leqslant t \leqslant 1\}$ 和 $\{N_2(t), 0 \leqslant t \leqslant 1\}$ 为独立泊松过程，且参数分别为 w_1 和 w_2. 此外，设 W_1 和 W_2 是参数为 Y 的独立指数随机变量，且 Y 本身为 $(0.02, 0.10)$ 上均匀分布的随机变量. 换言之，假设球员击出本垒打的次数服从一个参数为随机变量的泊松过程，且这一随机变量的分布的参数本身服从已知分布. 257

假设在前半赛季邦兹击出 25 次本垒打，格里非 18 次. 给出一种方法来估计他们个人在全赛季分别击出的本垒打均值.

求解　概括模型知，存在随机变量 Y，W_1，W_2 满足：

1. Y 服从 $(0.02, 0.10)$ 上的均匀分布.

2. 给定 $Y=y$，W_1 和 W_2 是参数为 y 的独立同分布的指数随机变量.

3. 给定 $W_1=w_1$ 和 $W_2=w_2$，$\{N_1(t)\}$ 和 $\{N_2(t)\}$ 是参数为 w_1 和 w_2 的独立泊松过程.

为求 $E[N_1(1)|N_1(0.5)=25, N_2(0.5)=18]$，首先加上 W_1 的条件.

$$E[N_1(1)|N_1(0.5)=25, N_2(0.5)=18, W_1]=25+0.5W_1$$

求在给定 $N_1(0.5)=25$ 和 $N_2(0.5)=18$ 下的条件期望，得

$$\begin{aligned}&E[N_1(1)|N_1(0.5)=25, N_2(0.5)=18]\\&\quad=25+0.5E[W_1|N_1(0.5)=25, N_2(0.5)=18]\end{aligned}$$

类似地有，

$$\begin{aligned}&E[N_2(1)|N_1(0.5)=25, N_2(0.5)=18]\\&\quad=18+0.5E[W_2|N_1(0.5)=25, N_2(0.5)=18]\end{aligned}$$

现利用吉布斯抽样估计这些条件期望. 首先，注意联合分布：对 $0.02<y<0.10$，$w_1>0$，$w_2>0$，

$$\begin{aligned}&f(y, w_1, w_2, N_1(0.5)=25, N_2(0.5)=18)\\&\quad=Cy^2\mathrm{e}^{-(w_1+w_2)y}\mathrm{e}^{-(w_1+w_2)/2}(w_1)^{25}(w_2)^{18}\end{aligned}$$

其中 C 不依赖于 y，w_1 和 w_2. 因此，对于 $0.02<y<0.10$，

$$f(y|w_1, w_2, N_1=25, N_2=18)=C_1y^2\mathrm{e}^{-(w_1+w_2)y}$$

上式说明在给定 $w_1, w_2, N_1=25$ 和 $N_2=18$ 下 Y 的条件分布为参数为 3 和 w_1+w_2 的伽玛随机变量在 0.02 到 0.10 间取值的条件分布. 另外，

[258]

$$f(w_1|y, w_2, N_1(0.5)=25, N_2(0.5)=18)=C_2\mathrm{e}^{-(y+1/2)w_1}(w_1)^{25}$$

由此可知，在给定 y，w_2，$N_1=25$ 和 $N_2=18$ 下 W_1 的条件分布是参数为 26 和 $y+\frac{1}{2}$ 的伽玛分布. 同理，在给定 y，w_1，$N_1=25$ 和 $N_2=18$ 下 W_2 的条件分布是参数为 19 和 $y+\frac{1}{2}$ 的伽玛分布.

于是，从 y, w_1 和 w_2 的值开始，其中 $0.02 < y < 0.10$，且 $w_i > 0$，吉布斯抽样如下进行：

1. 产生一个值在 0.02 与 0.10 之间的参数为 3 和 $w_1 + w_2$ 的伽玛随机变量的值，并令其为新的 y 值.
2. 产生参数为 26 和 $y + \frac{1}{2}$ 的一个伽玛随机变量的值，并令其为新的 w_1.
3. 产生参数为 19 和 $y + \frac{1}{2}$ 的一个伽玛随机变量的值，并令其为新的 w_2.
4. 回到步骤 1.

w_1 的平均值即为 $E[W_1|N_1(0.5) = 25, N_2(0.5) = 18]$ 的估计，w_2 的平均值即为 $E[W_2|N_1(0.5) = 25, N_2(0.5) = 18]$ 的估计. 前者的一半加上 25 即为我们对巴里 · 邦兹全年击出本垒打的平均数的估计，后者的一半加上 18 为肯 · 格里非将击出的本垒打的平均数估计.

应该注意到两球员的本垒打数是相关的，其相关性在于它们都依赖于随机变量 Y. 即，Y 的值（与很多因素有关，如这个赛季中使用棒球的平均灵活度或是全年中平均的天气情况）影响每位球员全年击出本垒打平均数的分布. 因此，由一个球员击出本垒打的次数的信息可以得到关于 Y 的概率信息，而 Y 值又影响其他球员本垒打数的分布. 这种模型，有共同的随机变量（在本例中为 Y）影响感兴趣随机变量的条件参数的分布，称为分层贝叶斯模型. □

在应用吉布斯抽样时，没有必要约束所有变量，而只约束其中之一即可. 如果抽样可能从联合分布中产生，则我们可以利用它们. 例如，设 $n = 3$ 且可从已知第三个随机变量的任意两个随机变量的条件分布产生抽样，则在每次重复中我们产生一个随机数 U，设 $I = \text{Int}(3U + 1)$，且从已知 X_I 当前值的 X_j, X_k, $j, k \neq I$ 的联合分布中产生此随机数.

259

例 10i 设 X_i, $i = 1, 2, 3, 4, 5$ 为独立指数随机变量，且 X_i 的均值为 i，并假设要利用模拟方法估计

$$\beta = P\left\{\prod_{i=1}^{5} X_i > 120 \middle| \sum_{i=1}^{5} X_i = 15\right\}$$

使用吉布斯抽样，随机选择坐标中的两个就可以完成此项任务. 首先，假设 X 和 Y 是参数分别为 λ 和 μ 的独立指数随机变量，其中 $\mu < \lambda$，并且我们按如下方法寻找在 $X + Y = a$ 下的 X 的条件分布：

$$\begin{aligned}
f_{X|X+Y}(x|a) &= C_1 f_{X,Y}(x, a - x), \quad & 0 < x < a \\
&= C_2 e^{-\lambda x} e^{-\mu(a-x)}, \quad & 0 < x < a \\
&= C_3 e^{-(\lambda-\mu)x}, \quad & 0 < x < a
\end{aligned}$$

此条件分布为一小于 a 条件下的参数为 $\lambda-\mu$ 的指数随机变量的条件分布.

利用这一结果，通过令初始状态 (x_1,x_2,x_3,x_4,x_5) 是和等于 15 的任意 5 个正数，我们可以估计 β. 现从 1, 2, 3, 4, 5 中随机选出两个数；假设挑选的两个数为 $I=2$ 和 $J=5$, 则在给定其他值下 X_2 和 X_5 的条件分布为在其和等于 $15-x_1-x_3-x_4$ 下两个均值分别为 2 和 5 的独立指数随机变量的条件分布. 但是，由前所述，要获得 X_2 和 X_5 的值，首先要产生一个小于 $15-x_1-x_3-x_4$ 且参数为 $\frac{1}{2}-\frac{1}{5}=\frac{3}{10}$ 的指数随机变量的值. 之后，令 x_2 等于这个值，且重设 x_5 使得 $\sum_{i=1}^{5}x_i=15$. 不断重复这一过程，用 $\prod_{i=1}^{5}x_i>120$ 的状态向量 $\boldsymbol{x}$ 的比例作为 β 的估计. □

例 10j　设做 n 次独立试验，每次可能结果为 $1,2,\cdots,r$ 的概率分别为 $p_1,p_2,\cdots,p_r$ 且 $\sum_{i=1}^{r}p_i=1$, 并令 X_i 为出现结果 i 的试验次数. 则随机变量 $X_1,\cdots,X_r$ 的联合分布称为多项分布，且在例 4g 中介绍了如何模拟这些变量. 现假设 $n>r$, 且要模拟 $X_1,\cdots,X_r$ 并使它们全部为正. 即，模拟试验使每种可能结果至少出现一次. 当此条件发生的概率非常小时，要如何有效完成模拟？

260

求解　首先，应该注意，假设只产生 $n-r$ 个试验的结果，之后，令 X_i 等于 1 加上 $n-r$ 个试验中出现结果 i 的试验数，这种做法是错误的. (即，除去所有试验结果均出现一次的 r 次试验后，对剩下的 $n-r$ 次试验进行模拟是行不通的.) 为解释错误的原因，令 $n=4$ 和 $r=2$. 则按上述方法，只有两次试验出现结果 1 的概率为 $2p(1-p)$, 其中 $p=p_1$. 而对于多项随机变量 X_1, X_2,

$$\begin{aligned}P\{X_1=2|X_1>0,X_2>0\}&=\frac{P\{X_1=2\}}{P\{X_1>0,X_2>0\}}\\&=\frac{P\{X_1=2\}}{1-P\{X_1=4\}-P\{X_2=4\}}\\&=\frac{\binom{4}{2}p^2(1-p)^2}{1-p^4-(1-p)^4}\end{aligned}$$

由于上式不等于 $2p(1-p)$ (代入 $p=1/2$), 故此方法行不通.

我们可使用吉布斯抽样产生一个有适当的极限概率的马氏链. 设初始状态为任一包括 r 个正数的随机向量，其和为 n, 并以如下方式改变状态. 如果状态为 $x_1,\cdots,x_r$, 则从下标 $1,\cdots,r$ 中随机选出两个，产生下一个状态. 如果选择下标为 i 和 j, 则令 $s=x_i+x_j$, 并从已知 $X_k=x_k$, $k\neq i,j$ 的它们的条件分布模拟生成 X_i 和 X_j. 由于以 $X_k=x_k$, $k\neq i,j$ 为条件，结果为 i 或 j 的试验总数为 s, 故出现结果 i 的试验次数服从一个参数为 $\left(s,\frac{p_i}{p_i+p_j}\right)$ 的二项分布，其可能值为 $1,\cdots,s-1$. 因此，可以使用离散逆变换方法模拟这样的一个随机变量；如果它的值为 v, 则下

一个状态除 x_i 和 x_j 有新的值 v 和 $s-v$ 外，与前面状态相同. 继续这一过程得到一系列极限分布为多项分布的状态，其中所有结果至少出现一次.

注 1 当考虑一列坐标并应用吉布斯抽样时 (如例 10i)，或者使用吉布斯抽样，没有约束所有变量，而只约束其中一个变量时 (如例 10j)，所获得的适当的极限分布函数均可由上述方法来检验. 注意到如果选择分布函数为 f 的初始状态，则对上述两种情况，下个状态也有分布函数 f，由此可证明前面的结论. 而且可以证明 f 满足 (10.1) 式，由唯一性知 f 为极限分布函数.

261

注 2 设在容易算出条件均值 $E[X_i|X_j, j \neq i]$ 的情况下，使用吉布斯抽样估计 $E[X_i]$. 则相对于使用 X_i 的一系列值的平均作为估计，使用条件期望的平均作为估计通常为更好的做法. 即，如果当前状态为 x，则通过模拟，取 $E[X_i|X_j = x_j, j \neq i]$ 而非 x_i 为估计. 类似地，如果要估计 $P\{X_i = x\}$，且 $P\{X_i = x|X_j, j \neq i\}$ 容易计算，则相对于状态向量中第 i 个元素为 x 的比例，这些量的平均通常为更好的估计.

注 3 吉布斯抽样表明给定其他 $X_j(j \neq i)$ 值下 X_i 的所有条件分布，可确定 $\boldsymbol{X}$ 的联合分布. □

10.4 模拟退火

令 $\mathcal{A}$ 为一有限向量集合且 $V(\boldsymbol{x})$ 是一定义在 $\boldsymbol{x} \in \mathcal{A}$ 上的非负函数. 假设现在我们感兴趣的是找到其最大值及至少一个达到其最大值的向量. 也就是说，令

$$V^* = \max_{\boldsymbol{x} \in \mathcal{A}} V(\boldsymbol{x})$$

及

$$\mathcal{M} = \{\boldsymbol{x} \in \mathcal{A} : V(\boldsymbol{x}) = V^*\}$$

我们感兴趣的是找到 V^* 及 $\mathcal{M}$ 中的一个元素. 下面将说明如何用本章中所介绍的方法来解决这个问题.

首先，令 $\lambda > 0$ 并考虑如下定义在 $\mathcal{A}$ 上的概率分布函数：

$$p_\lambda(\boldsymbol{x}) = \frac{\mathrm{e}^{\lambda V(\boldsymbol{x})}}{\sum\limits_{\boldsymbol{x} \in \mathcal{A}} \mathrm{e}^{\lambda V(\boldsymbol{x})}}$$

将上式的分子分母同时乘以 $\mathrm{e}^{-\lambda V^*}$，并令 $|\mathcal{M}|$ 表示 $\mathcal{M}$ 中的元素的个数，可得到

$$p_\lambda(\boldsymbol{x}) = \frac{\mathrm{e}^{\lambda(V(\boldsymbol{x}) - V^*)}}{|\mathcal{M}| + \sum\limits_{\boldsymbol{x} \notin \mathcal{M}} \mathrm{e}^{\lambda(V(\boldsymbol{x}) - V^*)}}$$

但是，由于当 $\boldsymbol{x} \notin \mathcal{M}$ 时，$V(\boldsymbol{x}) - V^* < 0$，故可得，当 $\lambda \to \infty$，

$$p_\lambda(\boldsymbol{x}) \to \frac{\delta(\boldsymbol{x}, \mathcal{M})}{|\mathcal{M}|}$$

262 其中若 $\boldsymbol{x} \in \mathcal{M}$，$\delta(\boldsymbol{x}, \mathcal{M}) = 1$，否则为零.

因此，如果我们令 λ 很大并生成一极限分布是 $p_\lambda(\boldsymbol{x})$ 的马氏链，则这个极限分布的绝大多数密度都集中在 $\mathcal{M}$ 中的点上. 定义这样的一个马氏链的有效方法是引入相邻向量的概念并使用 Hastings-Metropolis 算法. 例如，如果两个向量 $\boldsymbol{x} \in \mathcal{A}$ 和 $\boldsymbol{y} \in \mathcal{A}$ 仅在一个坐标上不同，或者其中的一个向量可以由另一个通过交换其两个分量而得到，则我们称它们是相邻的. 然后，如我们令从 $\boldsymbol{x}$ 出发的下一状态等可能地取其任意一个相邻向量，且若选择了 $\boldsymbol{y}$，那么下一个状态为 $\boldsymbol{y}$ 的概率为

$$\min\left\{1, \frac{\mathrm{e}^{\lambda V(\boldsymbol{y})}/|N(\boldsymbol{y})|}{\mathrm{e}^{\lambda V(\boldsymbol{x})}/|N(\boldsymbol{x})|}\right\}$$

或者仍然为 $\boldsymbol{x}$，这里 $|N(\boldsymbol{z})|$ 表示 $\boldsymbol{z}$ 的相邻向量的个数. 如果每个向量有同样数目的相邻向量（如果这个现在不成立，我们总还是能够通过扩大状态空间并令任意新状态的 V 值为 0 的方法使其成立），那么当状态为 $\boldsymbol{x}$ 时，随机选取其相邻向量 $\boldsymbol{y}$；如果 $V(\boldsymbol{y}) \geqslant V(\boldsymbol{x})$，那么马氏链转移到状态 $\boldsymbol{y}$，而如果 $V(\boldsymbol{y}) < V(\boldsymbol{x})$，则马氏链以概率 $\exp\{\lambda(V(\boldsymbol{y}) - V(\boldsymbol{x}))\}$ 转移到状态 $\boldsymbol{y}$ 或者仍然为 $\boldsymbol{x}$.

前面算法的一个缺点是，由于 λ 比较大，如果马氏链进入一个其 V 值比它的每一个相邻向量的 V 值都大的状态 $\boldsymbol{x}$，那么也许需要较长的一段时间才能使马氏链转移到一个不同的状态. 也就是说，鉴于极限分布需要一个较大的 λ 值才能使得其密度集中在 $\mathcal{M}$ 中的点上，在接近极限分布之前，这样的一个值特别地需要很大次数的转移. 第二个缺点是由于 $\boldsymbol{x}$ 的可能值只有限多个，且我们总是从理论上尝试每一个可能值，以致于有限步后自然会收敛，所以，整个收敛的概念看上去没有意义. 于是，我们在这里不从严格的数学角度出发来考虑前面所述方法，而更合理的是把它认为是一种探索性的方法，这样的话我们发现允许 λ 的值随时间变化是很有用的.

前面方法的一个常用的变形，也就是熟知的*模拟退火方法*，是如下运作的. 如果马氏链的第 n 个状态为 $\boldsymbol{x}$，那么随机选取其一相邻向量. 如果是 $\boldsymbol{y}$，那么下一个状态或者以概率

$$\min\left\{1, \frac{\mathrm{e}^{\lambda_n V(\boldsymbol{y})}/|N(\boldsymbol{y})|}{\mathrm{e}^{\lambda_n V(\boldsymbol{x})}/|N(\boldsymbol{x})|}\right\}$$

为 $\boldsymbol{y}$ 或者仍为 $\boldsymbol{x}$，其中 $\lambda_n, n \geqslant 1$ 是一组事先选定的集合，开始为小值（因此导致状态变化的次数很多）然后逐渐增加.

一个对计算很有帮助的 λ_n 的选取（并且是从数学上来说收敛更好）是令 $\lambda_n = C\ln(1+n)$，这里 $C>0$ 是任意固定的正常数 [1]. 接下来，如果我们生成 m 个连续的状态 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m$，则我们可以用 $\max_{i=1,\cdots,m} V(\boldsymbol{X}_i)$ 来估计 V^*，且如果最大值在 $\boldsymbol{X}_{i^*}$ 点取得，那么它就被当作是一 $\mathcal{M}$ 中的估计点. 263

例 10k　旅行商问题　一个版本的旅行商问题是商人从城市 0 出发，然后依次拜访所有的城市 $1,\cdots,r$. 一种选择是取 $1,\cdots,r$ 的一个排序 $x_1,\cdots,x_r$，然后按照此排序，他从城市 0 出发，到城市 x_1，然后到城市 x_2，依次类推. 如果我们假设无论何时他若从 i 直接到 j，他将能够获得一个非负的报酬 $v(i,j)$，那么使用 $\boldsymbol{x}=(x_1,\cdots,x_r)$ 将会获得回报

$$V(\boldsymbol{x}) = \sum_{i=1}^{r} v(x_{i-1},x_i),\ \text{其中}\ \ x_0 = 0$$

现假设如果一个排序是由另一个排序的两个坐标相交换而得到的，那么这两个排序被认为是相邻的，之后我们可以使用模拟退火的方法来估计最优路径. 也就是说，从任意的排序 $\boldsymbol{x}$ 开始且令 $\boldsymbol{X}_0=\boldsymbol{x}$. 只要第 $n(n\geqslant 0)$ 个状态（也就是，排序）被确定了，之后随机地生成其相邻排序 [在 $\binom{r}{2}$ 值中等可能地选取 i 和 j，且 $i\neq j, i,j=1,\cdots,r$，然后交换 $\boldsymbol{X}_n$ 的第 i 个和第 j 个元素的值]. 记生成的相邻排序为 $\boldsymbol{y}$. 如果 $V(\boldsymbol{y})\geqslant V(\boldsymbol{X}_n)$，则置 $\boldsymbol{X}_{n+1}=\boldsymbol{y}$. 否则，以概率 $(1+n)^{(V(\boldsymbol{y})-V(\boldsymbol{X}_n))}$ 置 $\boldsymbol{X}_{n+1}=\boldsymbol{y}$，或者令它等于 $\boldsymbol{X}_n$. [注意我们这里使用 $\lambda_n=\ln(1+n)$.] □

10.5　抽取重要再抽样的算法

抽取重要再抽样的算法，或者说 SIR，是一种通过模拟一个马氏链来生成一密度函数为

$$f(\boldsymbol{x}) = C_1 f_o(\boldsymbol{x})$$

的随机向量 $\boldsymbol{X}$ 的方法，其中 C_1 是需确定的常数，且马氏链的极限概率密度函数为

$$g(\boldsymbol{x}) = C_2 g_o(\boldsymbol{x})$$

其中 C_2 也是需确定的一个常数. 类似于筛选法，也就是首先生成密度为 g 的随机向量 $\boldsymbol{Y}$，然后如果 $\boldsymbol{Y}=\boldsymbol{y}$，则以概率 $f(\boldsymbol{y})/cg(\boldsymbol{y})$ 接受这个值，其中 c 是一选定的常数且满足对所有的 $\boldsymbol{x}$ 有 $f(\boldsymbol{x})/cg(\boldsymbol{x})\leqslant 1$. 如果不接受这个值，那么过程重新开始， 264

1. 见 Besag *et al.*, 1995；Diaconis and Holmes，1995.

且最终接受的值 $\boldsymbol{X}$ 有密度函数 f. 但是，由于 f 和 g 不再是完全确定的，故这种方法是不适用的.

SIR方法首先生成极限概率密度函数为 g 的马氏链的 m 个连续状态. 记这些状态为 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_m$. 现定义 "权" $w_i, i=1,\cdots,m$ 为

$$w_i = \frac{f_o(\boldsymbol{y}_i)}{g_o(\boldsymbol{y}_i)}$$

并生成随机变量 $\boldsymbol{X}$ 使得

$$P\{\boldsymbol{X}=\boldsymbol{y}_j\} = \frac{w_j}{\sum\limits_{i=1}^{m} w_i},\ j=1,\cdots,m$$

我们下面证明当 m 很大时，随机向量 $\boldsymbol{X}$ 的密度函数近似地为 f.

命题　当 $m\to\infty$ 时，通过 SIR 方法得到的随机向量 $\boldsymbol{X}$ 的分布收敛到 f.

证明　令 $\boldsymbol{Y}_i, i=1,\cdots,m$ 表示由极限密度函数为 g 的马氏链生成的 m 个随机向量，且用 $W_i=f_o(\boldsymbol{Y}_i)/g_o(\boldsymbol{Y}_i)$ 表示它们的权. 对于固定的向量集合 $\mathcal{A}$，若 $\boldsymbol{Y}_i\in\mathcal{A}$，则令 $\boldsymbol{I}_i=1$，否则为 0. 于是，

$$P\{\boldsymbol{X}\in\mathcal{A}|\boldsymbol{Y}_i, i=1,\cdots,m\} = \frac{\sum\limits_{i=1}^{m} I_i W_i}{\sum\limits_{i=1}^{m} W_i} \tag{10.5}$$

由马氏链的结果 (10.2) 可知，当 $m\to\infty$ 时，

$$\sum_{i=1}^{m} I_i W_i/m \to E_g[IW] = E_g[IW|I=1]P_g\{I=1\} = E_g[W|\boldsymbol{Y}\in\mathcal{A}]P_g\{\boldsymbol{Y}\in\mathcal{A}\}$$

及

$$\sum_{i=1}^{m} W_i/m \to E_g[W] = E_g[f_o(\boldsymbol{Y})/g_o(\boldsymbol{Y})] = \int \frac{f_o(\boldsymbol{y})}{g_o(\boldsymbol{y})} g(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = C_2/C_1$$

因此，将 (10.5) 的分子分母同时除以 m 可有

265

$$P\{\boldsymbol{X}\in\mathcal{A}|\boldsymbol{Y}_i, i=1,\cdots,m\} \to \frac{C_1}{C_2} E_g[W|\boldsymbol{Y}\in\mathcal{A}]P_g\{\boldsymbol{Y}\in\mathcal{A}\}$$

而

$$\begin{aligned}
\frac{C_1}{C_2} E_g[W|\boldsymbol{Y}\in\mathcal{A}]P_g\{\boldsymbol{Y}\in\mathcal{A}\} &= \frac{C_1}{C_2} E_g\left[\frac{f_o(\boldsymbol{Y})}{f_o(\boldsymbol{Y})}|\boldsymbol{Y}\in\mathcal{A}\right] P_g\{\boldsymbol{Y}\in\mathcal{A}\} \\
&= \int_{\boldsymbol{y}\in\mathcal{A}} \frac{f(\boldsymbol{y})}{g(\boldsymbol{y})} g(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\
&= \int_{\boldsymbol{y}\in\mathcal{A}} f(\boldsymbol{y})\mathrm{d}\boldsymbol{y}
\end{aligned}$$

因此，当 $m \to \infty$ 时，

$$P\{\boldsymbol{X} \in \mathcal{A} | \boldsymbol{Y}_i, i = 1, \cdots, m\} \to \int_{\boldsymbol{y} \in \mathcal{A}} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

由勒贝格控制收敛定理可得

$$P\{\boldsymbol{X} \in \mathcal{A}\} = E[P\{\boldsymbol{X} \in \mathcal{A} | \boldsymbol{Y}_i, i = 1, \cdots, m\}] \to \int_{\boldsymbol{y} \in \mathcal{A}} f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

由此结论得证. □

用抽取重要再抽样的算法来近似地生成具有密度函数 f 的一随机向量，它首先以一不同的联合密度函数来生成随机变量（就如在重要抽样中一样），然后从这些生成的值中重抽样获得随机向量.

假设对于某个函数 h，要估计 $E_f[h(\boldsymbol{X})]$. 我们可以首先生成一极限概率为 g 的马氏链的大量的连续状态. 如果这些状态为 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m$，那么看上去可以很自然地选取 k 个具有概率函数

$$P\{\boldsymbol{X} = \boldsymbol{y}_j\} = \frac{w_j}{\sum\limits_{i=1}^{m} w_i}, \quad j = 1, \cdots, m$$

的向量 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_k$，其中 k/m 很小，而 $w_i = f_o(\boldsymbol{y}_i)/g_o(\boldsymbol{y}_i)$，然后可用 $\sum_{i=1}^{k} h(\boldsymbol{X}_i)/k$ 作为估计量.

然而，一个更好的方法是不使用基于 k 个值的抽样集合的估计，而是用整个 m 个生成值 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m$. 下面我们说明

$$\frac{1}{\sum\limits_{i=1}^{m} w_i} \sum_{j=1}^{m} w_j h(\boldsymbol{y}_j)$$

是 $E_f[h(\boldsymbol{X})]$ 的一个比 $\sum_{i=1}^{k} h(\boldsymbol{X}_i)/k$ 更好的估计. 为证明这点，注意到

$$E[h(\boldsymbol{X}_i) | \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m] = \frac{1}{\sum\limits_{i=1}^{m} w_i} \sum_{j=1}^{m} w_j h(\boldsymbol{y}_j)$$

266

因此

$$E\left[\frac{1}{k} \sum_{i=1}^{k} h(\boldsymbol{X}_i) | \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\right] = \frac{1}{\sum\limits_{i=1}^{m} w_i} \sum_{j=1}^{m} w_j h(\boldsymbol{y}_j)$$

这就说明 $\sum_{j=1}^{m} h(\boldsymbol{y}_j) w_j / \sum_{i=1}^{m} w_i$ 与 $\sum_{i=1}^{k} h(\boldsymbol{X}_i)/k$ 有相同的均值，但前者方差更小.

使用从一个分布中生成的数据来收集另一个分布的信息在贝叶斯统计上特别有用.

例 101　假设 $\boldsymbol{X}$ 是一概率分布函数为一指定的未知参数向量 $\boldsymbol{\theta}$ 的随机向量. 例如, $\boldsymbol{X}$ 可以是一列独立同分布的正态随机变量且 $\boldsymbol{\theta}=(\theta_1,\theta_2)$, 其中 θ_1 和 θ_2 分别是这些随机变量的均值和方差. 令 $f(\boldsymbol{x}|\boldsymbol{\theta})$ 表示给定 $\boldsymbol{\theta}$ 下 $\boldsymbol{X}$ 的密度. 在经典统计中一般假设 $\boldsymbol{\theta}$ 为一未知的常数向量, 而在贝叶斯统计中我们将其假设为一个有密度函数 $p(\boldsymbol{\theta})$ 的随机变量, 称之为先验密度.

如果观测到 $\boldsymbol{X}$ 等于 $\boldsymbol{x}$, 那么 $\boldsymbol{\theta}$ 的条件, 或者说后验的密度就是

$$p(\boldsymbol{\theta}|\boldsymbol{x})=\frac{f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}(\boldsymbol{\theta})}$$

然而, 在许多情形下 $\int f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}(\boldsymbol{\theta})$ 是难以计算的, 因此, 不能够直接使用上式来研究后验分布.

一种研究后验分布性质的方法是首先通过先验密度 p 来生成随机向量 $\boldsymbol{\theta}$, 然后用这些生成的数据来收集后验密度 $p(\boldsymbol{\theta}|\boldsymbol{x})$ 的信息. 如果我们假设先验密度 $p(\boldsymbol{\theta})$ 是完全确定的且可直接用它生成随机变量, 那么我们就可以通过

$$f_o(\boldsymbol{\theta})=f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$
$$g(\boldsymbol{\theta})=g_o(\boldsymbol{\theta})=p(\boldsymbol{\theta})$$
$$w(\boldsymbol{\theta})=f(\boldsymbol{x}|\boldsymbol{\theta})$$

来使用SIR算法. 首先, 生成大量来自于先验密度 $p(\boldsymbol{\theta})$ 的随机向量, 记其值为 $\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_m$. 然后用估计量

267

$$\sum_{j=1}^{m}\alpha_j h(\boldsymbol{\theta}_j),\ \ \text{其中}\ \alpha_j=\frac{f(\boldsymbol{x}|\boldsymbol{\theta}_j)}{\sum\limits_{i=1}^{m}f(\boldsymbol{x}|\boldsymbol{\theta}_i)}$$

来估计形式为 $E[h(\boldsymbol{\theta})|\boldsymbol{x}]$ 的任意函数. 举例来说, 对任意的集合 $\mathcal{A}$, 我们可以使用

$$\sum_{j=1}^{m}\alpha_j I\{\boldsymbol{\theta}_j\in\mathcal{A}\}\ \ \text{来估计}\ P\{\boldsymbol{\theta}\in\mathcal{A}|\boldsymbol{x}\}$$

其中当 $\boldsymbol{\theta}_j\in\mathcal{A}$ 时, $I\{\boldsymbol{\theta}_j\in\mathcal{A}\}$ 为 1, 否则为 0.

在 $\boldsymbol{\theta}$ 的维数较小时, 我们可以利用来自先验分布的数据以及它们的权用图例方式研究其后验分布. 举例来说, 如果 $\boldsymbol{\theta}$ 是二维的, 那么我们可以将生成的值 $\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_m$ 表示在一个二维图上, 并将点的权考虑进去. 比如, 我们可以对每个点都画一个小圆点, 而在点 $\boldsymbol{\theta}_j$ 所画的圆点的面积与其权 $f(\boldsymbol{x}|\boldsymbol{\theta}_j)$ 成比例. 另一种方法

是所有的圆点都同样大小，但是点的灰暗度以一线性可加的方式依赖于它们的权.比如，如果 $m=3$，且 $\boldsymbol{\theta}_1=\boldsymbol{\theta}_2$，$f(\boldsymbol{x}|\boldsymbol{\theta}_3)=2f(\boldsymbol{x}|\boldsymbol{\theta}_1)$，那么在 $\boldsymbol{\theta}_1$ 所画的点的色彩应同 $\boldsymbol{\theta}_3$ 的一致.

如果先验密度 p 仅仅是一给定的常数，或者难以通过它生成随机向量，那么我们可以生成一具有极限密度 p 的马氏链，然后用前面介绍的方法继续来做. □

注 在例 10l 中所介绍的 $E[h(\boldsymbol{\theta})|\boldsymbol{x}]$ 估计也可以由重要抽样方法推导出来. 由于

$$p(\boldsymbol{\theta}|\boldsymbol{x})=\frac{f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{C}$$

其中

$$C=\int f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}=E[f(\boldsymbol{x}|\boldsymbol{\theta})]$$

我们有

$$\begin{aligned}E[h(\boldsymbol{\theta})|\boldsymbol{x}]&=\int\frac{h(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{x})}{p(\boldsymbol{\theta})}p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\\&=E\left[\frac{h(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{x})}{p(\boldsymbol{\theta})}\right]\\&=\frac{E[f(\boldsymbol{x}|\boldsymbol{\theta})h(\boldsymbol{\theta})]}{C}\\&=\frac{E[f(\boldsymbol{x}|\boldsymbol{\theta})h(\boldsymbol{\theta})]}{E[f(\boldsymbol{x}|\boldsymbol{\theta})]}\end{aligned}$$

268

由上述可知，其估计可取为

$$\frac{\sum_{j=1}^{m}f(\boldsymbol{x}|\boldsymbol{\theta}_j)h(\boldsymbol{\theta}_j)}{\sum_{i=1}^{m}f(\boldsymbol{x}|\boldsymbol{\theta}_i)}$$

这也就是例 10l 中给出的估计. □

习 题

1. 设 $\pi_j, j=1,\cdots,N$ 为一马氏链的平稳概率. 证明：如果 $P\{X_0=j\}=\pi_j, j=1,\cdots,N$，则

$$P\{X_n=j\}=\pi_j,\quad 对所有的 n,j$$

2. 设 $\boldsymbol{Q}$ 为对称转移概率矩阵，即 $q_{ij}=q_{ji}$，对所有 i,j. 设一马氏链满足：如果当前状态为 i，则产生的随机变量 X 的值满足 $P\{X=j\}=q_{ij}$；如果 $X=j$，则或以概率 $b_j/(b_i+b_j)$ 转移到状态 j，或者仍停留在状态 i，其中 b_j，$j=1,\cdots,N$ 为已知正数. 证明：所得马氏链是时间可逆的，且其极限概率为 $\pi_j=Cb_j$，$j=1,\cdots,N$.

3. 设 S 为所有元素为 0 或 1 的 $n \times n$ 矩阵 $\boldsymbol{A}$ 的集合. (因此, 集合 S 中有 2^{n^2} 个矩阵.) 矩阵 $\boldsymbol{A}$ 中的一对元素称为相邻, 如果 $|r-i|+|s-j|=1$. (例如, 与 $a_{2,2}$ 相邻的是 $a_{1,2}$, $a_{2,1}$, $a_{2,3}$ 和 $a_{3,2}$.) 设 $\mathcal{N}$ 为 $\boldsymbol{A}$ 中所有相邻元素对构成的集合, 矩阵 $\boldsymbol{A}$ 的 "Ising 能量" 定义为

$$H(\boldsymbol{A}) = -\sum_{\mathcal{N}} a_{i,j} a_{r,s}$$

其中和为所有相邻元素对的累加. 给出随机选取矩阵 $\boldsymbol{A}$ 的方法, 且其概率密度函数为

$$P(\boldsymbol{A}) = \frac{\exp\{-\lambda H(\boldsymbol{A})\}}{\sum\limits_{\boldsymbol{A}\in S} \exp\{-\lambda H(\boldsymbol{A})\}}, \quad \boldsymbol{A} \in S$$

其中 λ 为一已知正常数.

[提示: 定义矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相邻, 如果 $\boldsymbol{A}-\boldsymbol{B}$ 仅有一个非零元素.]

4. 考虑一包含 20 个部件的系统, 部件 i 以概率 $0.5+i/50$, $i=1,\cdots,20$ 起作用. 令 X 为起作用的部件数. 利用模拟方法估计条件概率密度函数 $P\{X=i|X \leqslant 5\}$, $i=1,2,3,4,5$.

269

5. 设随机变量 X 和 Y 均在区间 $(0,B)$ 上取值. 设在 $Y=y$ 下 X 的条件密度为

$$f(x|y) = C(y)\mathrm{e}^{-xy}, \quad 0 < x < B$$

及在 $X=x$ 下 Y 的条件密度为

$$f(y|x) = C(x)\mathrm{e}^{-xy}, \quad 0 < y < B$$

给出一种近似模拟向量 (X,Y) 的方法, 且利用模拟方法估计 (a) $E[X]$ 和 (b) $E[XY]$.

6. 给出一种产生 $(0,1)$ 上的 9 个均匀点的有效方法, 其中任意两个点间距离均大于 0.1.[可以证明如果 n 个点是独立的且服从 $(0,1)$ 上的均匀分布, 则它们中任意两个点间距离均大于 0.1 的概率为 $[1-(n-1)d]^n$, 其中 $0<d<1/(n-1)$.]

7. 在例 10d 中, 可以证明: 在 $m+1$ 名服务员上的顾客数的极限密度函数为

$$p(n_1,\cdots,n_m,n_{m+1}) = C\prod_{i=1}^{m+1} P_i(n_i), \quad \sum_{i=1}^{m+1} n_i = r$$

其中对 $i=1,\cdots,m+1$, $P_i(n)$, $n=0,\cdots,r$, 为一概率密度函数. 令 $\boldsymbol{e}_k$ 为一 $m+1$ 维向量, 其中第 k 个位置为 1, 其他为 0. 对于向量 $\boldsymbol{n}=(n_1,\cdots,n_{m+1})$, 令

$$q(\boldsymbol{n},\boldsymbol{n}-\boldsymbol{e}_i+\boldsymbol{e}_j) = \frac{I(n_i>0)}{(m+1)\sum\limits_{j=1}^{m+1} I(n_j>0)}$$

总之, q 为马氏链的转移概率矩阵, 在每一步随机地选取一个非空的服务员, 并将他的一位顾客随机地分配给另一个服务员. 使用上述 q 函数, 给出 Hastings-Metropolis 算法产生极限密度函数为 $p(n_1,\cdots,n_m,n_{m+1})$ 的马氏链.

8. 设 X_i, $i=1,2,3$ 为独立的均值是 1 的指数随机变量. 利用模拟方法估计

   (a) $E[X_1+2X_2+3X_3|X_1+2X_2+3X_3>15]$.

   (b) $E[X_1+2X_2+3X_3|X_1+2X_2+3X_3<1]$.

270

9. 从包含 n 个球的罐中随机抽取 m 个球，其中颜色类型 $i=1,\cdots,r$ 的球为 n_i 个，$\sum_{i=1}^{r} n_i = n$. 令 X_i 为抽取的球中颜色类型为 i 的球数. 给出一种方法模拟 $X_1,\cdots,X_r$，条件为所有 r 种颜色都被选出，且假设所有颜色都被选出的概率很小.

10. 假设 X,Y,Z 的联合密度函数为

$$f(x,y,z)=C\mathrm{e}^{-(x+y+z+axy+bxz+cyz)},\quad x>0,\ y>0,\ z>0$$

其中 a,b,c 为给定的非负常数，且 C 不依赖于 x,y,z. 解释我们如何来模拟向量 X,Y,Z 并模拟估计在 $a=b=c=1$ 情况下 $E[XYZ]$ 的值.

11. 假设对于随机变量 X,Y,N，有

$$P\{X=i,y\leqslant Y\leqslant y+\mathrm{d}y,N=n\}$$

$$\approx C\binom{n}{i}y^{i+\alpha-1}(1-y)^{ni+\beta-1}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\mathrm{d}y$$

其中 $i=0,\cdots,n$, $n=0,1,\cdots$, $y\geqslant 0$，并且 α,β,λ 为给定的常数. 在 $\alpha=2$, $\beta=3$, $\lambda=4$ 时，做模拟来估计 $E[X]$, $E[Y]$ 和 $E[N]$.

12. 用 SIR算法来生成 $1,2,\cdots,100$ 的一个排序，满足它的分布近似地为 $X_1,\cdots,X_{100}$ 在条件 $\sum_j jX_j>285\,000$ 下的一随机排序的分布.

13. 令 $\boldsymbol{X}^1,\boldsymbol{X}^2,\cdots,\boldsymbol{X}^n$ 为 $\mathcal{L}$ 中的随机点，其中 $\mathcal{L}$ 是以圆点为中心半径为 1 的圆. 假设对于某个 r, $0<r<1$，它们的密度函数为

$$f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=K\exp\{-\beta t(r:\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)\},\quad \boldsymbol{x}_i\in\mathcal{L},\ i=1,\cdots,n$$

其中 $t(r:\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$ 是 $\binom{n}{2}$ 对点 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$, $i\neq j$ 中互相之间距离小于 r 的数目，且 $0<\beta<\infty$. （注意 $\beta=\infty$ 对应如下情形：在任意两点之间的距离都不小于 r 的约束下 $\boldsymbol{X}^i$ 均匀地分布于圆上.）请说明我们如何利用 SIR 算法来近似地生成这些随机点. 如果 r 和 β 都很大，这个算法有效吗？

14. 生成 100 个随机数 $U_{0,k}, k=1,\cdots,10, U_{i,j}, i\neq j$, $i,j=1,\cdots,10$. 现考虑如下旅行商问题，商人从城市 0 出发必须根据 $1,\cdots,10$ 的某个排序依次访问每一个城市. 令 U_{ij} 为他直接从城市 i 到 j 所获得的报酬. 用模拟退火方法来估计商人可能获得的最大回报.

271

参考文献

[1] Aarts, E., and J. Korst, *Simulated Annealing and Boltzmann Machines.* Wiley, New York, 1989.

[2] Besag, J., *Towards Bayesian Image Analysis, J. Appl. Statistics*, **16**, 395-407, 1989.

[3] Besag, J., P. Green, D. Higdon, and K. Mengersen, *Bayesian Computation and Stochastic Systems (with Discussion), Statistical Sci.*, **10**, 3-67, 1995.

[4] Diaconis, P., and S. Holmes,*Three Examples of Monte-Carlo Markov Chains: At the Interface between Statistical Computing, Computer Science, and Statistical Mechanics, Discrete Probability and Algorithms* (D. Aldous, P. Diaconis, J. Spence, and J. M. Steele, eds.), pp.43-56. Springer-Verlag, 1995.

[5] Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. Smith, *Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling, J. Am. Statistical Assoc.*, **85**, 972-985, 1990.

[6] Gelfand, A. E., and A. F. Smith,*Sampling Based Approaches to Calculating Marginal Densities, J. Am. Statistical Assoc.*, **85**, 398-409, 1990.

[7] Gelman, A., and D. B. Rubin,*Inference from Iterative Simulation (with Discussion), Statistical Sci.*, **7**, 457-511, 1992.

[8] Geman, S., and D. Geman,*Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images,IEEE Trans. Pattern Anal. Machine Intelligence*, **6**, 721-724, 1984.

[9] Geyer, C. J.,*Practical Markov Chain Monte Carlo (with Discussion),Statistical Sci.*,**7**, 473-511, 1992.

[10] Gidas, B.,*Metropolis-type Monte Carlo Simulation Algorithms and Simulated Annealing,* in *Trends in Contemporary Probability* (J. L. Snell, ed.). CRC Press. Boca Raton, FL, 1995.

[11] Hajek, B.,*Cooling Schedules for Optimal Annealing, Math. Operations Res.*, **13**, 311-329, 1989.

[12] Hammersley, J. M., and D. C. Handscomb, *Monte Carlo Methods.* Methuen, London, 1965.

[13] Ripley, B., *Stochastic Simulation.* Wiley, New York, 1987.

[14] Rubin, D. R.,*Using the SIR Algorithm to Simulate Posterior Distributions,* in *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.), pp. 395-402. Oxford University Press, 1988.

[15] Rubinstein, R. R., *Monte Carlo Optimization, Simulation, and Sensitivity of Queueing Networks*. Wiley, New York, 1986.

[16] Sinclair, A., *Algorithms for Random Generation and Counting*. Birkhauser, Boston, 1993.

[17] Smith, A. F., and G. O. Roberts, *Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with Discussion)*, *J. Roy. Statistical Soc., Ser. B*, **55**, 3-23, 1993.

272

第11章　其他议题

引　言

在这一章里，我们探讨一些比前面章节考虑的问题更专业的话题. 在 11.1 节里，我们给出一种被称作别名方法的用于生成离散随机变量的高效技术. 在 11.2 节里，我们定义并给出模拟二维泊松过程的方法. 在 11.3 节，我们给出一个关于伯努利随机变量和的概率等式并说明利用它可在估计小概率问题上起到大量缩减方差的效果. 最后，11.4 节考虑如何有效地估计关于马氏链过程首达时的概率和期望.

11.1　用于生成离散随机变量的别名方法

在这一节里，我们研究一种生成离散随机变量的技术，这种方法尽管需要一些设置的时间，但是可以非常快速地实现.

现令 $\boldsymbol{P}, \boldsymbol{P}^{(k)}, \boldsymbol{Q}^{(k)}, k \leqslant n-1$ 表示在整数点 $1, 2, \cdots, n$ 取值的概率密度函数，也就是说，它们是和为 1 的非负的 n 维向量. 另外，向量 $\boldsymbol{P}^{(k)}$ 至多有 k 个非零的分量而每个 $\boldsymbol{Q}^{(k)}$ 有至多两个非零的分量. 我们来证明任何一个概率密度函数 $\boldsymbol{P}$ 可以表示为 $n-1$ 个概率密度函数 $\boldsymbol{Q}$（每一个都有至多两个非零的分量）的等权的加和. 也就是说，我们将证明，通过适当定义 $\boldsymbol{Q}^{(1)}, \cdots, \boldsymbol{Q}^{(n-1)}$，$\boldsymbol{P}$ 可以表示为

273

$$\boldsymbol{P} = \frac{1}{n-1} \sum_{k=1}^{n-1} \boldsymbol{Q}^{(k)} \tag{11.1}$$

为了给出我们的方法，需要如下一个简单的引理（其证明留作练习）.

引理　*令 $P = \{P_i, i = 1, \cdots, n\}$ 表示一个概率密度函数. 则*

(a) *存在一个 i $(1 \leqslant i \leqslant n)$，使得 $P_i < 1/(n-1)$；*

(b) *对于这个 i 存在一个 j $(j \neq i)$，使得 $P_i + P_j \geqslant 1/(n-1)$.*

在介绍获得 (11.1) 的通用方法之前，先用一个例子作说明.

例 11a　考虑三点分布 $\boldsymbol{P}$，其中 $P_1 = 7/16$，$P_2 = 1/2$，$P_3 = 1/16$. 我们首先来选择 i 和 j 使其满足前面引理中的条件. 因为 $P_3 < 1/2$ 且 $P_3 + P_2 \geqslant 1/2$，我们可取 $i = 3$ 及 $j = 2$. 现定义一个两点密度函数 $\boldsymbol{Q}^{(1)}$，将其所有的权重都置于 3 和 2 上使得 $\boldsymbol{P}$ 可表示为 $\boldsymbol{Q}^{(1)}$ 和另一个两点密度函数 $\boldsymbol{Q}^{(2)}$ 的等权加和. 另外，所有点

3 的密度包含在 $\boldsymbol{Q}^{(1)}$ 中. 由于有

$$P_j = \frac{1}{2}(Q_j^{(1)} + Q_j^{(2)}), \quad j = 1,2,3, \tag{11.2}$$

且假设 $Q_3^{(2)}$ 为零, 所以必须取

$$Q_3^{(1)} = 2P_3 = \frac{1}{8}, \quad Q_2^{(1)} = 1 - Q_3^{(1)} = \frac{7}{8}, \quad Q_1^{(1)} = 0$$

为了满足 (11.2), 则须令

$$Q_3^{(2)} = 0, \quad Q_2^{(2)} = 2P_2 - \frac{7}{8} = \frac{1}{8}, \quad Q_1^{(2)} = 2P_1 = \frac{7}{8}$$

至此, 在这个例子中得到了想要的表达式. 假设现考虑的原始分布是下面的四点密度函数:

$$P_1 = \frac{7}{16}, \quad P_2 = \frac{1}{4}, \quad P_3 = \frac{1}{8}, \quad P_4 = \frac{3}{16}$$

现有 $P_3 < 1/3$ 及 $P_3 + P_1 \geqslant 1/3$. 因为初始的两点密度函数 $\boldsymbol{Q}^{(1)}$ 将密度都置于点 3 和 1 上 (点 2 和 4 上没有权重). 所以最终的表达式中, $\boldsymbol{Q}^{(1)}$ 的权重为 1/3 并且其他的 $\boldsymbol{Q}^{(j)}, j = 2,3$ 对点 3 不给予密度, 必须有 274

$$\frac{1}{3}Q_3^{(1)} = P_3 = \frac{1}{8}$$

因此,

$$Q_3^{(1)} = \frac{3}{8}, \quad Q_1^{(1)} = 1 - \frac{3}{8} = \frac{5}{8}$$

并且, 有

$$\boldsymbol{P} = \frac{1}{3}\boldsymbol{Q}^{(1)} + \frac{2}{3}\boldsymbol{P}^{(3)}$$

这里 $\boldsymbol{P}^{(3)}$ 为了满足上式, 须为

$$\boldsymbol{P}_1^{(3)} = \frac{3}{2}\left(P_1 - \frac{1}{3}Q_1^{(1)}\right) = \frac{11}{32}$$
$$\boldsymbol{P}_2^{(3)} = \frac{3}{2}P_2 = \frac{3}{8}$$
$$\boldsymbol{P}_3^{(3)} = 0$$
$$\boldsymbol{P}_4^{(3)} = \frac{3}{2}P_4 = \frac{9}{32}$$

注意 $\boldsymbol{P}_3^{(3)}$ 在点 3 的密度为零. 现在可以把密度函数 $\boldsymbol{P}^{(3)}$ 表示为两点密度函数 $\boldsymbol{Q}^{(2)}$ 和 $\boldsymbol{Q}^{(3)}$ 的等权加和. 最后得到

$$\begin{aligned}\boldsymbol{P} &= \frac{1}{3}\boldsymbol{Q}^{(1)} + \frac{2}{3}\left(\frac{1}{2}\boldsymbol{Q}^{(2)} + \frac{1}{2}\boldsymbol{Q}^{(3)}\right) \\ &= \frac{1}{3}(\boldsymbol{Q}^{(1)} + \boldsymbol{Q}^{(2)} + \boldsymbol{Q}^{(3)})\end{aligned}$$

(剩余具体细节留给读者作为练习.) □

上面的例子概述了将一个 n 点的密度函数 $\boldsymbol{P}$ 表示为形如 (11.1) 的表达式的过程, 其中每个 $\boldsymbol{Q}^{(i)}$ 有至多两个非零的分量. 首先我们选择 i 和 j 来满足引理的条件. 注意到在表达式 (11.1) 中 $Q_i^{(k)} = 0, k = 2, \cdots, n-1$, 意味着

275

$$Q_i^{(1)} = (n-1)P_i \quad 并且 \quad Q_j^{(1)} = 1-(n-1)P_i.$$

因此我们定义密度函数 $\boldsymbol{Q}^{(1)}$, 使其密度集中于点 i 和 j 上, 并且包含 i 点的所有密度. 有

$$\boldsymbol{P} = \frac{1}{n-1}\boldsymbol{Q}^{(1)} + \frac{n-2}{n-1}\boldsymbol{P}^{(n-1)}, \tag{11.3}$$

其中 $\boldsymbol{P}^{(n-1)}$ 表示剩余的密度, 可知

$$\begin{aligned}&P_i^{(n-1)} = 0\\ &P_j^{(n-1)} = \frac{n-1}{n-2}\left(P_j - \frac{1}{n-1}Q_j^{(1)}\right) = \frac{n-1}{n-2}\left(P_i + P_j - \frac{1}{n-1}\right)\\ &P_k^{(n-1)} = \frac{n-1}{n-2}P_k, \quad k \neq i \text{ 或 } j\end{aligned}$$

我们可以容易地验证上面的表达式确实是一概率密度函数, 举例来说, $P_j^{(n-1)}$ 的非负性可由 $P_i + P_j \geqslant 1/(n-1)$ 得到.

对于在点 $(n-1)$ 的概率密度函数 $\boldsymbol{P}^{(n-1)}$, 重复上面的过程得到

$$\boldsymbol{P}^{(n-1)} = \frac{1}{n-2}\boldsymbol{Q}^{(2)} + \frac{n-3}{n-2}\boldsymbol{P}^{(n-2)}$$

并且由 (11.3) 我们有

$$\boldsymbol{P} = \frac{1}{n-1}\boldsymbol{Q}^{(1)} + \frac{1}{n-1}\boldsymbol{Q}^{(2)} + \frac{n-3}{n-1}\boldsymbol{P}^{(n-2)}$$

对于 $\boldsymbol{P}^{(n-2)}$ 重复这个过程, 并且一直继续直到最终获得

$$\boldsymbol{P} = \frac{1}{n-1}(\boldsymbol{Q}^{(1)} + \cdots + \boldsymbol{Q}^{(n-1)})$$

通过这种方法我们能够将 $\boldsymbol{P}$ 表示为 $n-1$ 个两点密度函数的等权加和. 现在可以通过首先生成一个随机整数 N, 其等可能地等于 $1,2,\cdots,n-1$ 中的一个, 来模拟来自于 $\boldsymbol{P}$ 的随机数. 如果得到的值 N 使得 $Q^{(N)}$ 将权仅置于点 i_N 和 j_N 上, 且生成的第二个随机数小于 $Q_{i_N}^{(N)}$, 则可令 X 等于 i_N, 否则, 令其等于 j_N. 随机变量 X 有密度函数 $\boldsymbol{P}$. 也就是说, 可通过如下过程生成来自 $\boldsymbol{P}$ 的随机变量.

276

1. 生成 U_1 且令 $N = 1 + \mathrm{Int}[(n-1)U_1]$.
2. 生成 U_2 且令

$$X = \begin{cases} i_N & \text{如果 } U_2 < Q_{i_N}^{(N)} \\ j_N & \text{其他情况} \end{cases}$$

注 1 上述方法之所以称为别名方法是因为通过对 $\boldsymbol{Q}$ 重新编号, 我们总可以使得对于每个 k 都有 $Q_k^{(k)} > 0$. (即, 我们可以安排第 k 个两点密度函数在 k 上的权为正.) 因此, 此过程需要模拟生成 N, 其等可能地为 $1,2,\cdots,n-1$ 中的一个, 那么如果 $N=k$, X 的值或者为 k, 或者为 k 的 "别名" (也就是, $\boldsymbol{Q}^{(k)}$ 给予权的其他值).

注 2 事实上, 在步骤 2 中我们不需要生成一个新的随机数. 因为 $N-1$ 是 $(n-1)U_1$ 的整数部分, 则小数部分 $(n-1)U_1-(N-1)$ 与 N_1 独立并且来自均匀分布 $(0,1)$. 因此, 在步骤 2 中我们可以用 $(n-1)U_1-(N-1)$ 而不用生成新的随机数 U_2.

11.2 生成二维泊松过程

一个由平面上的随机点组成的过程被称作有参数 $\lambda, \lambda > 0$ 的二维泊松过程, 如果

1. 出现在任意给定的面积为 A 的区域内点的数目服从均值为 λA 的泊松分布.
2. 出现在不相交的区域内的点的数目是独立的.

对于给定的平面上的点 $\mathbf{0}$, 现在我们来说明如何根据一个参数为 λ 的二维泊松过程来生成以 $\mathbf{0}$ 为中心半径为 r 的圆形区域内的随机点.

令 $C(a)$ 表示以 $\mathbf{0}$ 为圆心半径为 a 的圆, 并且注意到, 由条件 1, 发生于 $C(a)$ 中的点的数目服从均值为 $\lambda\pi a^2$ 的泊松分布. 令 $R_i, i \geqslant 1$ 表示从原点 $\mathbf{0}$ 到其第 i 个最近的点的距离 (如图 11-1).

则

$$P\{\pi R_1^2 > x\} = P\{R_1 > \sqrt{x/\pi}\}$$

$$= P\left\{在\ C\left(\sqrt{\frac{x}{\pi}}\right)\ 中没有点\right\}$$

277

$$= \mathrm{e}^{-\lambda x}$$

其中最后一个等式利用了 $C(\sqrt{x/\pi})$ 的面积为 x 这一结论. 并且, 用 $C(b)-C(a)$ 表示 $C(b)$ 和 $C(a)$ 之间的区域 $(a<b)$, 有

$$\begin{aligned}
&P\{\pi R_2^2 - \pi R_1^2 > x | R_1 = a\}\\
&= P\left\{R_2 > \sqrt{\frac{x+\pi R_1^2}{\pi}} | R_1 = a\right\}\\
&= P\left\{在\ C\left(\sqrt{\frac{x+\pi a^2}{\pi}}\right) - C(a)\ 中没有点 \mid R_1 = a\right\}\\
&= P\left\{在\ C\left(\sqrt{\frac{x+\pi a^2}{\pi}}\right) - C(a)\ 中没有点\right\} \quad 由条件\ 2\\
&= \mathrm{e}^{-\lambda x}.
\end{aligned}$$

事实上, 继续反复利用上面的论证, 我们可获得如下命题.

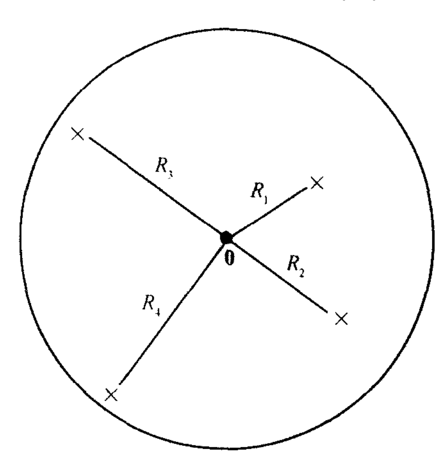


图 11-1 图 $C(a)$ 内点到原点的距离

命题 令 $R_0 = 0, \pi R_i^2 - \pi R_{i-1}^2, i \geqslant 1$ 是一列独立的服从参数为 λ 的指数分布的随机变量.

换言之, 为遇到一泊松点所需穿过的区域面积服从参数为 λ 的指数分布. 因此由对称性, 泊松点之间的夹角独立同分布于 $(0, 2\pi)$ 上的均匀分布, 因此我们可利用如下的算法来模拟以 $\mathbf{0}$ 为中心半径为 r 的圆形区域上的泊松过程.

278

1. 生成独立的参数为 λ 的指数随机变量，$X_1, X_2, \cdots$，直到

$$N = \min\{n : X_1 + \cdots + X_n > \pi r^2\}$$

时停止.

2. 如果 $N = 1$, 即在 $C(r)$ 中没有点，则停止；否则，对 $i = 1, \cdots, N-1$，令

$$R_i = \sqrt{\frac{X_1 + \cdots + X_i}{\pi}}$$

（即，$\pi R_i^2 = X_1 + \cdots + X_i$）.

3. 生成随机数 $U_1, \cdots, U_{N-1}$.

4. $N-1$ 个泊松点的极坐标为

$$(R_i, 2\pi U_i),\ i = 1, \cdots, N-1$$

上述算法可以被看作是一个以 0 为中心的圆随着半径从 0 变化为 r 不断地向外铺展的过程. 利用在遇到另外一个点之前所需要穿过的面积总是服从参数为 λ 的指数分布这一结论，接连地生成所遇点所在圆的半径. 这一向外铺展的技术同样可以被用来模拟生成在非圆形区域上的过程. 举例来说，考虑一非负函数 $f(x)$ 且假设我们感兴趣的是生成位于 x 轴及函数 f 之间从 0 到 T 的区域上的泊松过程（如图 11-2）. 为了达到此目的，考虑所经过的连续区域的面积，我们可以从左边的边缘开始垂直地向右边铺展. 特别地，如果 $X_1 < X_2 < \cdots$ 表示泊松点在 x 轴上的连续投影，则其形式与前面（$X_0 = 0$ 时）的完全一样：

$$\int_{X_{i-1}}^{X_i} f(x)\mathrm{d}x,\quad i = 1, \cdots, \text{为独立同分布的参数为 } \lambda \text{ 的指数分布.}$$

279

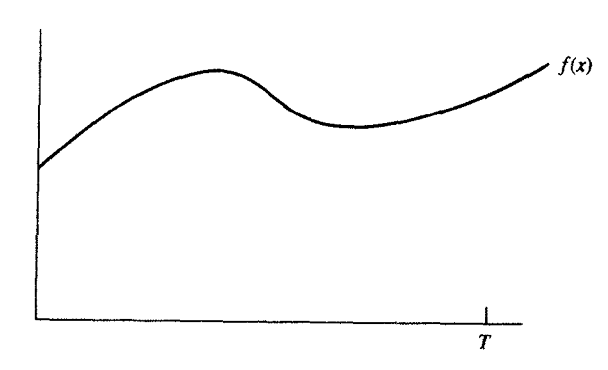


图 11-2 一般形式的二维泊松过程

因此，可以通过生成独立的参数为 λ 的指数随机变量 $W_1, W_2, \cdots$ 来模拟泊松点，直到

$$N = \min\left\{n : W_1 + \cdots + W_n > \int_0^T f(x)\mathrm{d}x\right\}$$

现通过如下等式来决定 $X_1, \cdots, X_{N-1}$：

$$\int_0^{X_1} f(x)\mathrm{d}x = W_1$$
$$\int_{X_1}^{X_2} f(x)\mathrm{d}x = W_2$$
$$\vdots$$
$$\int_{X_{N-2}}^{X_{N-1}} f(x)\mathrm{d}x = W_{N-1}$$

因为横坐标为 X_i 的点在 y 轴上的投影显然服从 $(0, f(X_i))$ 上的均匀分布，因此如果生成随机数 $U_1, \cdots, U_{N-1}$，则模拟得到的泊松点的直角坐标为 $(X_i, U_i f(X_i)), i = 1, \cdots, N-1$.

当 f 足够规则以便我们利用上面等式能够有效地解出 X_i 值的时候，上述方法是最为有用的. 例如，如果 $f(x) = c$（因此区域就为一矩形），我们可以将 X_i 表示为

$$X_i = \frac{W_1 + \cdots + W_i}{c}$$

则泊松点即为

$$(X_i, cU_i), \quad i = 1, \cdots, N-1$$

11.3 关于一个伯努利随机变量和的恒等式的应用模拟

令 $X_1, \cdots, X_m$ 为伯努利随机变量，且

$$P\{X_i = 1\} = \lambda_i = 1 - P\{X_i = 0\}$$

另外，令 $S = \sum_{i=1}^m X_i$，及 $\lambda = E[S] = \sum_{i=1}^m \lambda_i$. 令 R 为任意一随机变量，并假设 I 与 $R, X_1, \cdots, X_m$ 独立，且满足

$$P\{I = i\} = 1/m, \ i = 1, \cdots, m$$

280 即，I 是 $1, \cdots, m$ 上的离散均匀分布且与其他的随机变量独立.

下面的恒等式是本节所有结果的关键.

命题

(a) $P\{I=i|X_I=1\}=\lambda_i/\lambda$

(b) $E[SR]=\lambda E[R|X_I=1]$

(c) $P\{S>0\}=\lambda E[\frac{1}{S}|X_I=1]$

证明 为证明 (a), 注意到

$$P\{I=i|X_I=1\}=\frac{P\{X_I=1|I=i\}P\{I=i\}}{\sum\limits_i P\{X_I=1|I=i\}P\{I=i\}}$$

现有,

$$\begin{aligned}P\{X_I=1|I=i\}&=P\{X_i=1|I=i\}\\&=P\{X_i=1\} \quad \text{由独立性}\\&=\lambda_i\end{aligned}$$

由此完成 (a) 的证明. (b) 的证明如下:

$$\begin{aligned}E[SR]&=E\left[R\sum_i X_i\right]=\sum_i E[RX_i]\\&=\sum_i\{E[RX_i|X_i=1]\lambda_i+E[RX_i|X_i=0](1-\lambda_i)\}\\&=\sum_i \lambda_i E[R|X_i=1]\end{aligned} \tag{11.4}$$

而且,

$$\begin{aligned}E[R|X_I=1]&=\sum_i E[R|X_I=1,I=i]P\{I=i|X_I=1\}\\&=\sum_i E[R|X_i=1,I=i]\lambda_i/\lambda \quad \text{由(a)}\\&=\sum_i E[R|X_i=1]\lambda_i/\lambda\end{aligned} \tag{11.5}$$

综合 (11.4) 和 (11.5) 可证明 (b). 281

为了证明 (c), 定义 R 为: 如果 $S=0$ 则 $R=0$, 若 $S>0$ 则 $R=1/S$. 那么

$$E[SR]=P\{S>0\} \quad 并且 \quad E[R|X_I=1]=E\left[\frac{1}{S}\middle|X_I=1\right]$$

则 (c) 可由 (b) 直接得到. □

下面，我们利用此引理来估计 (a) 一个系统的失效概率及 (b) 指定模式在给定时间内发生的概率.

例 11b 考虑例 8b 中的模型，也就是一个由 n 个独立的元件组成的系统，并假设我们想要估计系统失效的概率，而此概率非常小. 现在，对任意的例 8b 中考虑的模型形式，总存在唯一的一个集合族 $\{C_1, \cdots, C_m\}$，且这些集合不相互包含，使得当且仅当至少其中的一个集合中的元件都失效时此系统失效. 这些集合被称作此系统的最小割集.

若元件 j 失效，则令 $Y_j = 1, j = 1, \cdots, n$，否则令其为 0. 并且令 $q_j = P\{Y_j = 1\}$ 表示元件 j 失效的概率. 现对 $i = 1, \cdots, m$，令

$$X_i = \prod_{j \in C_i} Y_j$$

也就是说，X_i 是 C_i 中所有元件失效这个事件的指示变量. 如果我们令 $S = \sum_i X_i$，则系统失效的概率 θ 就是

$$\theta = P\{S > 0\}$$

下面来介绍如何利用命题有效地估计 θ.

首先，令 $\lambda_i = E[X_i] = \prod_{j \in C_i} q_j$ 及 $\lambda = \sum_i \lambda_i$. 模拟生成随机变量 J，其值以概率 λ_i/λ 等于 i，$i = 1, \cdots, m$. [由命题的 (a) 可知 J 的分布与给定 $X_I = 1$ 时 I 的条件分布相同.] 则对 $i \in C_j$，令 $Y_i = 1$. 并模拟其他 $i \notin C_j$ 的 Y_i 值，令它们以概率 q_i 等于 1，以概率 $1 - q_i$ 等于 0. 令 S^* 表示所得的最小割集的数目，并注意到 $S^* \geqslant 1$. 由命题的 (c) 可知 λ/S^* 是 θ 的一个无偏估计. 因为 $S^* \geqslant 1$，则有

$$0 \leqslant \lambda/S^* \leqslant \lambda$$

且因此当最小割集的数量的均值 λ 非常小的时候，估计量 λ/S^* 会有非常小的方差.

282

例如，考虑一个 3/5 系统，也就是五个元件中至少有三个失效系统会失效，并且假设每个元件独立地以概率 q 失效. 对这样的一个系统，最小割集将是 $\binom{5}{3} = 10$ 个容量为 3 的子集. 因为所有的元件失效的概率是相同的，则 I 的值不会起任何作用. 因此，前面所介绍的估计可以通过假定元件 1,2 和 3 都失效并生成另外两个元件的状态的方法来决定. 于是，通过考虑元件 4 和 5 失效的数目，且由于 $\lambda = 10q^3$，

我们可得此估计的分布为

$$P\{\lambda/S^* = 10q^3\} = (1-q)^2$$
$$P\{\lambda/S^* = 10q^3/4\} = 2q(1-q)$$
$$P\{\lambda/S^* = q^3\} = q^2$$

又 $p = 1-q$, 我们有

$$\begin{aligned}\text{Var}(\lambda/S^*) &= E[(\lambda/S^*)^2] - (E[\lambda/S^*])^2 \\ &= 100q^6[p^2 + pq/8 + q^2/100 - (p^2 + pq/2 + q^2/10)^2]\end{aligned}$$

下表给出了某些不同的 q 值，所对应的 θ 的值以及 $\text{Var}(R)$ 与估计量 λ/S^* 的方差的比值，其中 $\text{Var}(R) = \theta(1-\theta)$ 是原始估计量的方差.

q	θ	$\text{Var}(R)/\text{Var}(\lambda/S^*)$
0.001	9.985×10^{-9}	8.896×10^{10}
0.01	9.851×10^{-6}	8 958 905
0.1	0.008 56	957.72
0.2	0.057 92	62.59
0.3	0.163 08	12.29

因此，对于小的 q, $\text{Var}(\lambda/S^*)$ 大约与 θ^2 同阶，而 $\text{Var}(R) \approx \theta$. □

例 11c 等待一个模式 令 $Y_i, i \geqslant 1$ 是一列独立同分布的离散随机变量，分布函数为 $P_j = P\{Y_i = j\}$. 令 $i_1, \cdots, i_k$ 为这些随机变量的可能取值的一个固定序列，并定义

$$N = \min\{i : i \geqslant k, Y_{i-j} = i_{k-j}, j = 0, 1, \cdots, k-1\}$$

也就是说，N 是模式 $i_1, \cdots, i_k$ 第一次发生的时间. 我们感兴趣的是当 θ 很小时，如何用模拟来估计 $\theta = P\{N \leqslant n\}$. 然而通常的模拟估计是通过如下方式获得的：生成一列随机变量直到模式发生或者直到时间 n 仍没有发生（在前一种情况下令估计为 1 而后者为 0）. 我们将利用前面的命题来获得更有效的估计. [283]

首先令

$$X_i = 1 \quad \text{如果 } Y_i = i_k, Y_{i-1} = i_{k-1}, \cdots, Y_{i-k+1} = i_1$$

否则为 0. 换句话说，如果模式在时刻 i 发生，$X_i = 1$（不必须为首次）. 令

$$S = \sum_{i=k}^{n} X_i$$

表示到时刻 n 为止，模式发生的次数，并且注意到

$$\theta = P\{N \leqslant n\} = P\{S > 0\}$$

因为对 $k \leqslant i \leqslant n$ 有

$$\lambda_i = P\{X_i = 1\} = P_{i_1} P_{i_2} \cdots P_{i_k} \equiv p$$

则利用命题可得

$$\theta = (n-k+1)pE\left[\frac{1}{S}\middle| X_I = 1\right]$$

其中 I 与 Y_j 独立且等可能地取 $k, \cdots, n$ 中的任意值. 因此，我们为估计 θ，首先模拟一等可能取 $k, \cdots, n$ 中任意值的随机变量 J，并令

$$Y_J = i_k, \quad Y_{J-1} = i_{k-1}, \cdots, \quad Y_{J-k+1} = i_1$$

然后根据分布函数 P 来模拟其他 $n-k$ 个 Y_i，并令 S^* 为模式发生的次数. 则 θ 的模拟估计量为

$$\hat{\theta} = \frac{(n-k+1)p}{S^*}$$

284 对于小的 $(n-k+1)p$ 来说，上式是一个非常有效的估计. □

11.4 估计马氏链首达时的分布及均值

我们用一个状态为非负整数的马氏链描述如下的从一个状态转移到另一状态的过程：只要现在处于状态 i，则不管前面处于何状态，下一个状态为 j 的概率就为 $P_{i,j}$，其中 $P_{i,j}, i \geqslant 0, j \geqslant 0$ 是给定的非负数，且满足

$$\sum_j P_{i,j} = 1, \quad i \geqslant 0$$

因此，若 X_k 表示此过程的第 k 个状态，则有

$$P\{X_{k+1} = j | X_k = i, X_{k-1}, \cdots, X_0\} = P_{i,j}$$

令 S 表示马氏链的一状态子集，且假设对一给定的初始状态 $X_0 = 0 \notin S$，我们感兴趣的是此马氏链首次进入 S 中的一个状态的时刻的分布. 特别地，我们想要估计

$$p = P\{T > n\}$$

其中

$$T = \min\{k : X_k \in S\}$$

原始模拟方法是首先生成序列 $X_1, \cdots, X_n$，即生成分布函数为

$$P\{X_1 = j | X_0 = 0\} = P_{0,j}, \quad j \geqslant 0$$

的随机变量 X_1. 如果生成的 X_1 的值为 i，那么 X_2 就通过下面的分布函数来生成

$$P\{X_2 = j | X_1 = i\} = P_{i,j}, \quad j \geqslant 0$$

以此类推. 然后，令

$$I = \begin{cases} 1 & \text{如果对所有的 } i = 1, \cdots, n,\ X_i \notin S \\ 0 & \text{其他情况} \end{cases}$$

重复整个过程多次，获得 I 的平均值就是 p 的原始模拟估计量. 285

为了获得比 I 有更小的方差的无偏估计，对 $i \notin S$，令

$$\alpha(i) = \sum_{k \notin S} P_{i,k}$$

为从状态 i 出发而下一个状态不在 S 中的概率. 以 X_1 是否属于 S 为条件，可有

$$\begin{aligned} E[I] &= E[I | X_1 \in S](1 - \alpha(0)) + E[I | X_1 \notin S]\alpha(0) \\ &= E[I | X_1 \notin S]\alpha(0). \end{aligned}$$

因此，根据 $X_1 \notin S$ 时 X_1 的条件分布来生成初始状态，由分层抽样的结果我们可获得一改进的估计量. 也就是说，如果我们由如下分布函数来生成 X_1：

$$P\{X_1 = j | X_0 = 0, X_1 \notin S\} = \frac{P_{0,j}}{\sum_{r \notin S} P_{0,r}}, \ j \notin S \tag{11.6}$$

并根据马氏链的转移概率来生成后面的状态，则我们会获得改进的估计量 $I\alpha(0)$. 然而，我们可以不通过马氏链转移概率来生成后面状态，而继续应用前面的方法来进一步缩小估计的方差. 也就是说，如果从分布函数 (11.6) 获得的 X_1 的值为 k，那么我们由给定 X_2 不属于 S 的条件分布函数来生成 X_2，即通过如下的分布函数来生成 X_2：

$$P\{X_2 = j | X_1 = k, X_2 \notin S\} = \frac{P_{k,j}}{\sum_{r \notin S} P_{k,r}}, \ j \notin S$$

继续重复上述方法即可获得 p 的估计量，记其为 $\hat{p}$，

$$\hat{p}=\prod_{i=1}^{n}\alpha(X_{i-1})$$

其中 $X_i, i>0$ 是在给定 X_{i-1} 的值及 $X_i\notin S$ 下的条件分布来生成的.

例 11d　估计多元尾部概率　令 $Y_1,\cdots,Y_n$ 为服从指定联合分布的随机变量. 假设对于每一个 $i=1,\cdots,n$，在给定 $Y_1,\cdots,Y_{i-1}$ 下 Y_i 的条件分布既可通过计算得到也可以由模拟获得. 现假设要估计

286

$$p=P\{Y_1>a_1,\cdots,Y_n>a_n\}$$

其中 $a_i, i=1,\cdots,n$ 为给定值. 考虑马氏链的第 j 个状态为向量 $(Y_1,\cdots,Y_j)$，我们可以利用前面所介绍的方法来估计 p. 令

$$S=\{(y_1,\cdots,y_j): j\leqslant n,\ y_k\leqslant a_k \text{ 对某些 } k\leqslant j\}$$

因此，若 T 为直到 S 中的一个状态发生时所需转移的次数，那么

$$T>n \Longleftrightarrow Y_1>a_1,\cdots,Y_n>a_n$$

所以，我们可如下估计 p:

1. 在 $Y_1>a_1$ 的条件下生成 Y_1；记生成的值为 y_1.

2. 对 $i=2$ 到 $n-1$，在 $Y_k=y_k, k=1,\cdots,i-1, Y_i>a_i$ 条件下生成 Y_i；记生成的值为 y_i. 由此过程所得到的 p 的估计为

$$\hat{p}=P\{Y_1>a_1\}\prod_{i=2}^{n}P\{Y_i>a_i|Y_1=y_1,\cdots,Y_{i-1}=y_{i-1}\}$$

下面以估计二元正态随机变量的尾部概率为例来说明上述方法. 如果随机变量 Y_1 和 Y_2 的联合密度函数为

$$f(y_1,y_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2-2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}$$

则称之为服从二元正态分布. 可以证明，Y_1 服从均值为 μ_1 方差为 σ_1^2 的正态分布；Y_2 服从均值为 μ_2 方差为 σ_2^2 的正态分布；并且 ρ 为它们之间的相关系数. 进一步，在给定 $Y_1=y_1$ 下 Y_2 的条件分布为正态分布，其均值为

$$\mu_2(y_1)\equiv E[Y_2|Y_1=y_1]=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(y_1-\mu_1)$$

方差为

$$\sigma_2^2(y_1) \equiv \mathrm{Var}(Y_2|Y_1=y_1) = \sigma_2^2(1-\rho^2)$$

因此，我们可以通过模拟大于 a_1 的 Y_1 来估计

$$p = P\{Y_1 > a_1, Y_2 > a_2\}$$

287

（关于如何有效地在一正态随机变量大于某常数条件下模拟生成其值，可参见 8.7 节）. 若模拟值为 y_1，则 p 的估计为

$$\begin{aligned}\hat{p} &= P\{Y_1 > a_1\}P\{Y_2 > a_2|Y_1 = y_1\}\\ &= \overline{\Phi}\left(\frac{a_1-\mu_1}{\sigma_1}\right)\overline{\Phi}\left(\frac{a_2-\mu_2(y_1)}{\sigma_2(y_1)}\right)\end{aligned}$$

其中 $\overline{\Phi}(x)$ 为一标准正态随机变量大于 x 的概率.

因为 $P\{Y_2 > a_2|Y_1 = y_1\}$ 关于 y_1 单调（若 $\rho > 0$ 则单调增，若 $\rho < 0$ 则单调减），所以我们可以将生成的 Y_1 的值作为一控制变量. 然而，为了达到此目的，我们首先必须计算其均值 $E[Y_1|Y_1 > a_1]$. 我们注意到，对于一个标准正态随机变量 Z，有

$$E[Z|Z>a] = \frac{1}{\sqrt{2\pi}\,\overline{\Phi}(a)}\int_a^\infty x\mathrm{e}^{-x^2/2}\mathrm{d}x = \frac{1}{\sqrt{2\pi}\,\overline{\Phi}(a)}\mathrm{e}^{-a^2/2}$$

因此，记

$$Y_1 = \mu_1 + \sigma_1 Z$$

则有

$$Y_1 > a_1 \Longleftrightarrow Z > \frac{a_1-\mu_1}{\sigma_1}$$

所以，

$$\begin{aligned}E[Y_1|Y_1>a_1] &= E\left[\mu_1+\sigma_1 Z|Z>\frac{a_1-\mu_1}{\sigma_1}\right]\\ &= \mu_1 + \frac{\sigma_1}{\sqrt{2\pi}\,\overline{\Phi}(\frac{a_1-\mu_1}{\sigma_1})}\mathrm{e}^{-\left(\frac{a_1-\mu_1}{\sigma_1}\right)^2/2}\end{aligned}$$

□

假设我们想估计直到马氏链进入 S 中的一个状态所需的平均时间 $E[T]$. 为了获得 $E[T]$，我们首先来说明 $E[T]$ 可以表示成概率 $P\{T > j\}$ 的和.

为了说明这点，现令 I_j 为事件 $T > j$ 发生的示性函数. 也即是说，

$$I_j = \begin{cases} 1 & 若\ T > j \\ 0 & 其他情况\end{cases}$$

288

因为

$$\sum_{j=0}^{\infty} I_j = \sum_{j=0}^{T-1} I_j + \sum_{j=T}^{\infty} I_j = T$$

则取期望后，可以得到

$$E[T] = E\left[\sum_{j=0}^{\infty} I_j\right] = \sum_{j=0}^{\infty} E[I_j] = \sum_{j=0}^{\infty} P\{T > j\}$$

为估计 $P\{T > j\}$，选取一个较大的数 n，模拟生成序列 $X_1, \cdots, X_n$(生成 X_i 的方法同前面介绍的一样，即以给定 X_{i-1} 的取值和事件 $X_i \notin S$ 为条件来生成 X_i). 在模拟生成这些量之后，利用马氏链的转移概率来模拟 $X_{n+1}, X_{n+2}, \cdots$，直到有落在 S 中的状态出现为止．（也就是说，以生成的 X_{i-1} 的值和事件 $X_i \notin S$ 为条件来生成 X_i，$i > n$.）如果 $X_m \in S$ 是模拟得到的最后的值，那么这一次模拟得到的 $P\{T > j\}$ 的估计 e_j 就为

$$e_j = \begin{cases} 1 & 若\ j = 0 \\ \prod_{i=1}^{j} \alpha(X_{i-1}) & 若\ 1 \leqslant j \leqslant n \\ \prod_{i=1}^{n} \alpha(X_{i-1}) & 若\ n < j < m \\ 0 & 若\ j \geqslant m \end{cases}$$

那么 $E[T]$ 的估计就为 $\sum_{j \geqslant 0} e_j$；多次重复这样的模拟，取其平均就得到了最终的估计.

11.5　过去耦合法

考虑一状态为 $1, \cdots, m$ 的不可约马氏链，且其转移概率为 $P_{i,j}$. 假设我们想要模拟的随机变量的分布就是这个马氏链的平稳分布（相关的定义可见 10.1 节）. 注意到，在 10.1 节中，我们通过任意选择一个初始状态并在一个固定的长时期内模拟所要的马氏链，取最后的状态作为生成的随机变量值. 在这一节中，我们给出一种方法来生成**严格**地服从平稳分布的随机变量.

从理论上来说，如果我们从时刻 $-\infty$ 以任意状态开始生成马氏链，则时刻 0 的状态将服从平稳分布. 假想我们就这么来做的话，并假定不同人在这些时间点上来生成下一个状态. 因此，若在时间点 $-n$ 的状态 $X(-n)$ 为 i，那么人 $-n$ 将会生成以

概率 $P_{i,j}$ 取值 j 的随机变量, 这个生成的值就是时间点 $-(n-1)$ 的状态. 现在假设人 -1 要提前来生成他的随机变量. 因为他不知道在时刻 -1 的状态将是什么, 则他会生成一个随机变量序列 $N_{-1}(i), i=1,\cdots,m$. 其中 $N_{-1}(i)$ 表示如果 $X(-1)=i$ 时的下一个状态, 其值以概率 $P_{i,j}$ 等于 j, $j=1,\cdots,m$. 如果现在得到 $X(-1)=i$, 那么人 -1 就会生成时刻 0 的状态为

$$S_{-1}(i)=N_{-1}(i),\quad i=1,\cdots,m$$

[也就是说, 当在时刻 -1 的状态为 i 时, $S_{-1}(i)$ 为在时刻 0 的状态.]

现假设人 -2 听说人 -1 提早做其模拟, 则她也决定早做. 于是, 她生成了一列随机变量 $N_{-2}(i), i=1,\cdots,m$, 其中 $N_{-2}(i)$ 以概率 $P_{i,j}, j=1,\cdots,m$ 等于 j. 即, 如果告诉她 $X(-2)=i$, 她将报告 $X(-1)=N_{-2}(i)$. 综合这点及前面人 -1 所生成的可知, 如果 $X(-2)=i$, 那么在时刻 0 模拟的状态就为:

$$S_{-2}(i)=S_{-1}(N_{-2}(i)),\quad i=1,\cdots,m$$

继续前面的这种过程, 假设人 -3 生成了一列随机变量 $N_{-3}(i), i=1,\cdots,m$, 这里 $N_{-3}(i)$ 为 $X(-3)=i$ 时所生成的下一个状态的值. 因此, 如果 $X(-3)=i$, 那么时刻 0 的模拟值就会为

$$S_{-3}(i)=S_{-2}(N_{-3}(i)),\quad i=1,\cdots,m$$

现在假设我们一直继续这个过程, 那么将会获得模拟函数:

$$S_{-1}(i), S_{-2}(i), S_{-3}(i),\cdots\quad i=1,\cdots,m$$

从前往后来看, 我们将会在某一个时刻, 比如说 $-r$, 得到其模拟函数 $S_{-r}(i)$ 为一常函数. 也就是说, 对于所有的状态 $i=1,\cdots,m$ 我们将有 $S_{-r}(i)$ 等于某个状态 j. 这也就意味着无论我们得到的从时刻 $-\infty$ 到 $-r$ 的模拟值为多少, 我们可以肯定的是在 0 时刻模拟值为 j. 那么 j 就可以认为是分布严格为马氏链平稳分布的随机变量的模拟值.

例 11e 考虑一状态为 1, 2, 3 的马氏链并假设模拟产生了

$$N_{-1}(i)=\begin{cases}3 & 若\ i=1\\ 2 & 若\ i=2\\ 2 & 若\ i=3\end{cases}$$

290

及

$$N_{-2}(i)=\begin{cases}1 & 若\ i=1\\ 3 & 若\ i=2\\ 1 & 若\ i=3\end{cases}$$

那么

$$S_{-2}(i)=\begin{cases}3 & 若\ i=1\\ 2 & 若\ i=2\\ 3 & 若\ i=3\end{cases}$$

如果

$$N_{-3}(i)=\begin{cases}3 & 若\ i=1\\ 1 & 若\ i=2\\ 1 & 若\ i=3\end{cases}$$

那么

$$S_{-3}(i)=\begin{cases}3 & 若\ i=1\\ 3 & 若\ i=2\\ 3 & 若\ i=3\end{cases}$$

因此，无论时刻 -3 的状态为何值，时刻 0 的状态都为 3. □

注 本节中所介绍的这种生成服从马氏链平稳分布的随机变量的方法称为过去耦合法.

习 题

1. 给出生成参数为 $(5, 0.4)$ 的二项分布的别名方法.
2. 解释在别名方法中，我们如何为 $\boldsymbol{Q}^{(k)}$ 编号使得 k 为 $\boldsymbol{Q}^{(k)}$ 赋权的两个点之一.
3. 完成例 11a 的细节.
4. 编写一个程序来生成落在半径为 R 中的圆域内的二维泊松随机点，并在 $\lambda=1$ 和 $R=5$ 时运行此程序，将获得的点画出来.

291

5. 用模拟来估计图 8-1 中当元件 1, 2 和 3 都以概率 0.05 失效而元件 4, 5 以概率 0.01 失效时，桥结构失效的概率. 并且，比较你所得的估计和原始估计的方差.
6. 利用模拟来估计在抛掷一均匀硬币 100 次的过程中连续出现 10 次正面的概率. 并比较你所得的估计和原始估计的方差.
7. 利用模拟来估计在抛掷一硬币 20 次的过程中出现模式“正反反正反”的概率，其中，第 i 次抛掷出现正面的概率为 $(i+10)/40, i=1,\cdots,20$，且假设各次抛掷独立.
8. 解释在例 11b 中为何使用对偶变量可以进一步缩小估计的方差.
9. 确定在习题 5 中使用对偶变量所能获得的额外的方差减小量.
10. 令 X_i 是均值为 $\lambda_i, i=1,\cdots,m$ 的伯努利随机变量，并令 $S=\sum_i a_i X_i$，其中 a_i 为正的常数. 设 R 为一任意的随机变量，且令 I 与其他的变量独立并有分布函数 $P\{I=i\}=a_i/\sum_i a_i, i=1,\cdots,m$. 记 $\lambda=\sum_i a_i\lambda_i$.

(a) 求 $P\{I=i|X_I=1\}$.

(b) 证明 $E[SR]=\lambda E[R|X_I=1]$.

(c) 证明 $P\{S>x\}=\lambda E[I(S>x)/S|X_I=1]$，其中，如果 $S>x$，则 $I(S>x)$ 为 1，否则等于 0.

11. 假设现有 n 个元件，其中元件 j 独立地以概率 $P_j, j=1,\cdots,n$ 工作. 假定我们现想做 m 次实验. 然而，为了做第 i 次实验，在集合 $A_i, i=1,\cdots,m$ 中的所有元件都必须正常工作. 如果实验 i 可以完成，那么我们就会获得一个数值 a_i. 假设最终的返回值为所有 m 次实验返回值的和，我们现在感兴趣的是估计这个返回值大于 x 的概率. 假设这个概率很小，请给出一个有效的估计方法.

12. 令 $Z_1, Z_2,\cdots$ 表示一列标准正态随机变量. 记

$$M_n=\frac{Z_n+Z_{n-1}+Z_{n-2}+Z_{n-3}}{4},\quad n\geqslant 4$$

并定义 N 为

$$N=\min\left\{n:|M_n|>\frac{3}{2}\right\}$$

292

我们想要用模拟来估计 $E[N]$. 请给出使用原始估计方法和 11.4 节中的方法所获得的估计的方差.（上面所考虑的问题称作移动平均控制图，被用来决定一服从标准正态分布的过程何时发生飘移.）

参考文献

[1] Kronmal, R. A., and A. V. Peterson, Jr., *On the Alias Method for Generating Random Variables from a Discrete Distribution. Am. Statistician*, **33**, 214-218, 1979.

[2] Peterson, A. V., Jr., and R. A. Kronmal, *On Mixture Methods for the Computer Generation of Random Variables. Am. Statistician*, **36**, 184-191, 1982.

[3] Ross, S. M., *A New Simulation Estimator of System Reliability, J. Appl. Mathematics Stochastic Anal.*, **7**, 331-336, 1994.

[4] Ross, S. M., and J. G. Shanthikumar, *Pricing Exotic Options, Monotonicity in Volatility and Efficient Simulation, Probability Eng. Informational Sci.*, **14**, 317-326, 2000.

293

索　引

索引中页码为英文原书页码, 与书中页边标注的页码一致.

D

G

H

I

K

L

M

N

O

P

R

T

U

V

W